(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 March 2001 (22.03.2001)

PCT

(10) International Publication Number
WO 01/19993 A2

(51) International Patent Classification[7]: C12N 15/13, C07K 14/35, A61K 38/00, 39/00, A61P 31/04, C12N 9/00

(21) International Application Number: PCT/US00/25512

(22) International Filing Date:
18 September 2000 (18.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/154,322 17 September 1999 (17.09.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/154,322 (CIP)
Filed on 17 September 1999 (17.09.1999)

(71) Applicants *(for all designated States except US)*: **UNIVERSITY OF MARYLAND, BALTIMORE** [US/US]; 520 West Lombard Avenue, Baltimore, MD 21201 (US). **UNITED STATES GOVERNMENT, as represented by DEPARTMENT OF VETERANS AFFAIRS** [US/US]; 810 Vermont Avenue, N.W., Washington, DC 20420 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **TRUCKSIS, Michele** [US/US]; 5072 Jericho Road, Columbia, MD 21044 (US).

(74) Agents: **ZELANO, Anthony, J.** et al.; Millen, White, Zelano & Branigan, P.C., Arlington Courthouse Plaza 1, 2200 Clarendon Boulevard, Suite 1400, Arlington, VA 22201 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/19993 A2

(54) Title: VIRULENCE GENES OF *M. MARINUM* AND *M. TUBERCULOSIS*

(57) Abstract: Methods for identifying, isolating and mutagenizing virulence genes of mycobacteria, *e.g., M. marinum* and *M. tuberculosis*, are described. Also described are isolated virulence genes and fragments of them, isolated gene products and fragments of them, avirulent bacteria in which one or more virulence genes are mutagenized, attenuated vaccines containing such mutant bacteria, and methods to elicit an immune response in a host, using such mutant bacteria.

VIRULENCE GENES OF *M. MARINUM* and *M. TUBERCULOSIS*

Background of the Invention

Mycobacteria are bacterial organisms which are implicated in diseases such as, *e.g.*, tuberculosis. It would be desirable to provide means for treating or preventing conditions caused by such mycobacteria, *e.g.*, by immunization.

Description of the Invention

This invention relates, *e.g.*, to virulence genes of mycobacteria. The invention provides methods to identify and isolate virulence genes of, for example, *Mycobacterium marinum*, a fish bacterium, and *Mycobacterium tuberculosis*, the primary etiologic agent of human tuberculosis. The invention also provides methods to mutagenize such virulence genes, thereby allowing the generation and isolation of avirulent mycobacteria. The invention also relates to isolated virulence genes and variants and fragments thereof; to isolated virulence gene products and variants and fragments thereof; to mutant, avirulent, bacteria; to attenuated vaccines comprising the mutant bacteria; and to methods to elicit an immune response in a host, using such mutant bacteria.

One embodiment of the invention is a method for identifying a virulence gene of *M. marinum*, comprising

a) mutagenizing *M. marinum* bacteria by introducing into said bacteria a plasmid which comprises a tagged (*e.g.*, signature-tagged) transposon, whereby the transposon integrates into and disrupts a gene in the bacteria,

b) introducing said mutagenized bacteria into a host susceptible to infection thereof (*e.g.*, a goldfish),

c) identifying a mutagenized bacterium which comprises a tagged transposon and which exhibits reduced viability in the host, compared to other mutagenized or (non-mutagenized) *M. marinum* bacteria,

d) cloning and/or sequencing (characterizing) a nucleic acid sequence which flanks the integrated transposon in said identified mutagenized bacterium, and

e) identifying a wild type *M. marinum* gene which comprises at least a portion of said flanking sequence.

Of course, the above method can be carried out using one or more of the steps, in any order, effective to achieve the intended purpose.

Another embodiment is a method for identifying a virulence gene of *M. tuberculosis*, comprising identifying an *M. marinum* virulence gene as described above, and further comprising,

comparing said flanking nucleic acid sequence to a databank of *M. tuberculosis* nucleic acid sequences, and/or comparing the sequences of peptides which are coded for by said flanking sequences to a known *M. tuberculosis* protein database, and

identifying an *M. tuberculosis* gene which comprises a sequence that is substantially identical to said flanking sequences and/or polypeptides encoded by them. In other embodiments, the degree of identity can be less than substantially identical, *e.g.*, about 35-50%, or about 50-70%, or about 70-90%.

Another embodiment is a method for isolating a mutagenized *M. marinum* bacterium which exhibits reduced virulence in a host susceptible to infection thereof compared to a non-mutagenized *M. marinum* bacterium, comprising integrating a tagged (*e.g.*, signature-tagged) transposon into the DNA of a *M. marinum* bacterium in a manner effective to produced reduced virulence, and isolating said mutagenized bacterium.

Another embodiment is an avirulent *M. marinum* bacterium in which one or more genes comprising a nucleic acid of SEQ ID NOs: 4, 6, 8, 10, 11, 13, 17, 21, 23, 25, 27, 29, 31, 35, 39, 41, 43 or 44 are mutated. Another embodiment is a pharmaceutical composition or an attenuated vaccine comprising such an avirulent *M. marinum* bacterium and a pharmaceutically acceptable carrier.

Another embodiment is an avirulent *M. tuberculosis* bacterium in which one or more virulence genes identified as described above are mutated. Another embodiment is an avirulent *M. tuberculosis* bacterium in which one or more of the genes encoding proteins Rv0822c, CY20G9.23 (Rv0497), the pks family, including *e.g.*, ppsE (Rv2935), psk6 (Rv0405), pks9 (Rv1664), pks8 (Rv1662), pks1 (Rv2946c), and pks002c, Rv3511, O08381 (Rv0357c), Rv3775, Rv3137, Rv2348c, Rv3860, mbtB (Rv2383c), Rv2181, Rv1954c, Rv0987, Rv3268, Rv2610c, nrp (pir E70751, Rv0101), mbtE (Rv2380c), Rv0236c or smc (Rv2922c) are mutated. Another embodiment is a pharmaceutical composition or an attenuated vaccine comprising one or more of the above avirulent *M.*

tuberculosis bacteria (*e.g.*, an *M. tuberculosis* strain constructed with one or more mutations in one or more of the above virulence genes) and a pharmaceutically acceptable carrier.

Another embodiment is an isolated nucleic acid of *M. marinum* comprising an oligonucleotide of SEQ ID NOs: 4, 6, 8, 10, 11, 13, 17, 21, 23, 25, 27, 29, 31, 35, 39, 41, 43 or 44, or a variant or fragment thereof. Another embodiment is a nucleic acid which is complementary to at least a portion of said isolated *M. marinum* nucleic acid, or which can hybridize to at least a portion of said isolated *M. marinum* nucleic acid under selected (*e.g.*, high) stringency conditions. In other embodiments, the isolated *M. marinum* nucleic acid is a gene; or the isolated *M. marinum* nucleic acid or fragments thereof are cloned into, and/or expressed in, an expression vector.

Another embodiment is an isolated nucleic acid of *M. tuberculosis*, comprising a virulence gene identified as above, or a variant or fragment thereof. Another embodiment is a nucleic acid which is complementary to at least a portion of said isolated *M. tuberculosis* nucleic acid, or which can hybridize to at least a portion of said isolated *M. tuberculosis* nucleic acid under selected (*e.g.*, high) stringency conditions. In other embodiments, the isolated *M. tuberculosis* nucleic acid or fragments thereof are cloned into, and/or expressed in, an expression vector.

Another embodiment is a method to elicit an immune response in a fish, comprising introducing into the fish an avirulent *M. marinum* bacterium made (*e.g.*, isolated, constructed) as described above. Another embodiment is a method to elicit an immune response in a human or non-human animal (*e.g.*, domestic or farm animal, such as a cow) host, comprising introducing into said host an avirulent, *M. tuberculosis* bacterium, in which one or more virulence genes identified as described above are mutated. Another embodiment is a method to elicit an immune response in a human host, comprising introducing into such host an avirulent *M. tuberculosis* bacterium in which one or more of the genes encoding proteins Rv0822c, CY20G9.23, the pks family of proteins, Rv3511, 008381, Rv3775, Rv3137, Rv2348c, Rv3860, mbtB, Rv2181, Rv1954c, Rv0987, Rv3268, Rv2610c, nrp (pir E70751), mbtE, Rv0236c or smc is mutated.

A wide variety of *Mycobacteria* species can be used in the invention. In a most preferred embodiment, the bacterium is *Mycobacterium marinum* (*M. marinum*), which causes fish tuberculosis, as well as, in humans, skin infection or localized nodular and ulcerated lesions (mariner's tuberculosis) on the extremities and, in immunocompromised patients, systemic disease; *Mycobacterium tuberculosis* (*M. tuberculosis)*, the primary etiologic agent for tuberculosis (TB) in man; or *Mycobacterium bovis* (*M. bovis*), which causes human or bovine tuberculosis. Other species of *Mycobacterium* which can be used in the invention include, *e.g.*, *M. bovis BCG*, *M. africanum*, *M. leprae*, *M. microti*, *M. smegmatis*, *M. vaccae*, *M. ulcerans*, *M. haemophilum*, *M. fortuitum*, *M. chelonae*, and others.

The term "virulent" in the context of mycobacteria refers to a bacterium or strain of bacteria that replicates within a host cell or animal within the mycobacterium host range at a rate which is detrimental to the cell or animal, or that induces a host response which is detrimental. More particularly, virulent mycobacteria persist longer in a host than avirulent bacteria. Virulent mycobateria are typically disease producing; and infection leads to various disease states including fulminant disease in the lung, disseminated systemic milliary tuberculosis, tuberculosis meningitis, and/or tuberculosis abscesses of various tissues. Infection by virulent mycobacteria often results in death of the host organism.

By contrast, the term "avirulent," as used herein, refers to a bacterium or strain of bacteria that does not replicate within a host cell or animal within its host range; replicates at a rate which is not significantly detrimental to the cell or animal; and/or does not induce a detrimental host response. An avirulent (*e.g.*, attenuated, non-pathogenic) strain is incapable of inducing a full suite of symptoms of the disease that is normally associated with its virulent pathogenic counterpart. Avirulent bacteria exhibit a reduced ability, or an inability, to survive in a host, but not all bacteria which exhibit such an impaired ability to survive in a host are avirulent. For example, in a simultaneous *in vivo* test of several mutant bacteria, certain mutants which are unable to compete with other mutants may not, when tested in the presence of the other strains, replicate efficiently or survive in the host; however, such bacteria, when tested individually, may prove to be virulent. An avirulent bacterium can contain one or more mutations in one or more virulence genes.

A "virulence gene" encodes a gene product ("virulence factor, virulence determinant") which contributes, directly or indirectly, to infection (*e.g.*, attachment, invasion, transport into the cell, replication, etc.) and/or to tissue destruction and/or disease. A virulence gene can code for or modify, *e.g.*, an adhesion molecule or other molecule which aids in the attachment to or invasion of a host cell; a toxin (*e.g.*, a secreted factor which can cause lysis or damage of a host cell -- for example, a small molecule such as a polyketide, or an enzyme such as a phospholipase, lipase, esterase or protease); a factor required for efficient secretion of such a toxin; a factor involved in intracellular multiplication or growth; a factor involved in resistance to host defenses; a factor which can stimulate a host cell to produce an inflammatory product or cytokine that can amplify tissue damage in a host; or a factor which regulates the production and/or activity of a virulence factor. Also included are certain functions which resemble "housekeeping" functions, *e.g.*, functions which allow bacteria to provide nutrients that are limiting in a host, such as factors which aid in the acquisition of iron, or certain enzymes of purine or pyrimidine biosynthesis. For a review of some of the putative or suspected virulence determinants of *Mycobacterium tuberculosis*, see Quinn *et al* (1996). *Curr. Top. Microbiol. Immunol.* 215, 131-156.

By a "host" for a bacterium is meant an organism, or a cell or tissue of an organism, which can be infected by the bacterium and which exhibits consequences of that infection. For example, *Mycobacterium marinum* can infect and cause symptoms in the frog (*Rana pipiens*) or in any of about 150 fresh-water or salt-water species of fish. In an especially preferred embodiment, the host for *Mycobacterium marinum* is the goldfish, *Carassius auratus*. Well-established animal models for *M. tuberculosis* include, *e.g.*, guinea pig, mouse, rabbit and monkey; and many natural hosts exist for that bacterium, including large animals such as the elephant. Many other bacteria/host combinations are possible. See, *e.g.*, B. Bloom, ed., (1994). *Tuberculosis: Pathogenesis, Protection, and Control*, ASM Press, Washington, D.C. Chapter 11, for a discussion of tuberculosis in wild and domestic animals.

A system in which goldfish are infected by *M. marinum* (the "goldfish model") offers a number of advantages for experimental studies. For example, *M. marinum* has a generation time of only 4 hours (as compared, *e.g.*, to the greater than 20 hour generation

time of *M. tuberculosis*), and studies with *M. marinum* can be carried out in a Biosafety Level 2 facility (whereas a Biosafety Level 3 facility is required, *e.g.*, for studies with *M. tuberculosis*). *M. marinum* can serve as an appropriate surrogate model for the study of *M. tuberculosis*. *M. marinum* and the *M. tuberculosis* complex have been shown to be closely related by, *e.g.*, DNA hybridization and 16S rRNA gene sequence analysis (see, *e.g.*, Tønjum *et al* (1998). *J. of Clinical Microbiology* 36, 918-925). The disease progression and symptoms of fish infected with *M. marinum* mimic those of humans infected with *M. tuberculosis*: in both types of hosts, organs in all parts of the body can be infected; both bacteria replicate within macrophages and reside in an endosomal compartment which is nonacidic and does not fuse with the lysosomal compartment; and both bacteria readily kill macrophages.

Examples 1B and 1C show, *e.g.*, that the pathology in the goldfish model parallels that of human tuberculosis. Depending on the dose of *M. marinum* organisms which is inoculated into a fish, acute or chronic disease is elicited. The pathology of the acute disease includes severe peritonitis and necrosis with all animals dying within 17 days of infection. The pathology of the chronic disease includes progressive granuloma formation. Granulomas with different histopathological features (necrotizing, non-necrotizing and caseous) are seen in the experimentally infected goldfish, which is consistent with the granuloma types seen in naturally infected animals and parallels the types of granulomas found in human tuberculosis. Isolation of *M. marinum* from fish tissue is possible throughout the course of the experiment presented in Example 1 (up to 16 weeks) indicating, as in human tuberculosis, the persistence of the organisms in the host. Example 2 shows that the goldfish model can be used to distinguish virulent and avirulent forms of *M. marinum*. Further disclosure of how to make the goldfish model, and how to use it, *e.g.*, to characterize molecular pathogenesis, can be found, *e.g.*, in Talaat A.M. *et al* (1998). *Infection and Immunity* 66, 2938-2942.

As an initial step in isolating virulence mutants, bacteria, *e.g.*, *M. marinum*, can be mutated by any of a variety of routine procedures which are well-known in the art, *e.g.*, exposure to chemical agents, irradiation, genetic engineering, transposon mutagenesis, or the like. As used in this application, the term a "mutation" means any change (in comparison with the appropriate parental strain) in the DNA sequence of an organism, *e.g.*,

a single (or multiple) base change, insertion, deletion, inversion, translocation, duplication, or the like. A mutation can be polar or non-polar, a frameshift or in phase. Preferably, in particular when a mutated bacterium is used as part of a treatment regimen or a vaccine, the mutation is substantially incapable of reverting to the wild type.

In a most preferred embodiment, mutagenesis is carried out by a transposon mutagenesis system that carries sequence-specific tags, sometimes known as signature-tagged mutagenesis (STM). The unique tag sequence allows differentiation of individual mutants among an inoculum pool of mutants. The STM protocol permits the screening of a large number of mutants using a small number of animals. This method was developed by Hensel *et al* (Hensel *et al* (1995). *Science* 269, 400-403; U.S. Pat. No. 5,876,931 to Holden). Variations of the method and procedures for using it to isolate bacterial virulence mutants are also disclosed in, *e.g.*, Shea *et al* (1996). *Proc. Natl. Acad. Sci.* 93, 2593-2597; Mei *et al* (1997). *Mol. Microbiol.* 26, 399-407; Schwan *et al* (1998). *Infec. Immun.* 66, 567-572; and Chiang *et al* (1998). *Mol. Microbiol.* 27, 797-805. Example 3 shows the use of the STM system for the mutagenesis of *M. marinum*.

Any of a variety of methods can be used to generate a bank of plasmids carrying unique signature-tagged transposons. A most preferred embodiment is shown in Example 3A. Here, 96 independent, non-cross-hybridizing, signature-tagged transposons, each of which is hybridization- and amplification-efficient, are cloned into a mycobacteria suicide vector which carries a selectable marker. Many variants of such vectors, carrying any of a variety of selectable markers, can be used, of course. In example 3A, the marker is a kanamycin-resistance gene.

To generate a mutant mycobacterium library, plasmids from a master plasmid collection are introduced individually (*e.g.*, separately) into mycobacteria, preferably *M. marinum*, by any of a variety of routine, art-recognized techniques (*e.g.*, phage transduction, shooting a "gene gun," electroporation, or other conventional techniques). In a most preferred embodiment, as shown in Example 3C, plasmids are introduced into *M. marinum* by electroporation. Any desired number of transformed bacteria can be selected from each transformation. In Example 3C, ninety-six transformations are performed, one with each of the 96 master plasmids; and ten independent transformants are selected from each transformation, to yield a library of 960 transformants. As Example

3B shows, the transposons integrate randomly into the *M. marinum* chromosome. In the ideal circumstance, each integrated transposon disrupts a different gene, or a different portion thereof, to create a library of, in this example, 960 differently mutagenized bacteria.

Pools of mutagenized bacteria, each of which can be detected independently by virtue of its unique signature tag, are introduced into an appropriate host, *e.g.*, a goldfish (an "input pool"). Bacteria may be introduced into an animal by any route, *e.g.*, orally, intraperitoneally, intravenously or intranasally; for fish, the preferred routes of administration are oral or, most preferably, intraperitoneal. It may be useful to compare, *e.g.*, virulence genes identified by oral administration to those identified by intraperitoneal administration, as some genes may be required to establish infection by one route but not by the other. Bacteria are left in the host for a suitable length of time, which is a function of both the microorganism and the host. A method for optimization of some of the infection parameters for the *M. marinum*/goldfish system is shown, *e.g.*, in Examples 1 and 2.

Assays are performed to determine whether the bacteria are able to survive in the host during the period of infection. Any of a variety of such assays can be used, *e.g.*, subtractive hybridization, differential display, or the like. In a most preferred embodiment, as shown in Example 4A, after an optimized period of infection by a pool of *M. marinum* mutants, fish are sacrificed and one or more internal organs, *e.g.*, spleen, liver, kidney, peritoneum, heart, pancreas, or other organs evident to one of skill in the art, are cultured to isolate the mutant bacteria which were able to survive in the fish, defined as the output pool. A hybridization protocol to identify mutants present in the input and output pools is described in Example 4A. Mutants which are present in the input pool, but which cannot be detected after a predetermined time of infection has elapsed in the output pool, are candidates for avirulent mutants, *i.e.*, mutants which are unable to infect, replicate and/or cause damage, in a particular cell type or tissue.

In order to confirm that an *M. marinum* mutant is avirulent, each putative virulence mutant can be re-examined individually, *e.g.*, in the goldfish model. In a preferred embodiment, the median survival time (MST) of goldfish infected with a lethal dose (about $5x10^8$ cfu) of a putative virulence mutant can be determined, and those mutants which

allow goldfish to survive longer than fish inoculated with an equivalent dose of wild type organisms are categorized as putative virulence mutants. Many other types of screening assays can be used, including Competitive Indices, histopathology examinations of one or more of the organs described above, colony counts in organ homogenates, and analysis of the ability of a mutant to induce granuloma formation. Representative protocols for each of these methods are described, *e.g.*, in Example 4B. In addition to confirming the existence of a virulence mutant, data collected on each mutant can yield clues to the pathogenesis pathways of *M. marinum* in the goldfish model. Methods to show that Koch's postulates have been fulfilled (proving that a postulated virulence gene is responsible for disease symptoms) are routine; one such method is presented in Example 8.

Alternative approaches to the STM technique can be used to identify avirulent *M. marinum* mutants. For example, one can screen a library of *M. marinum* cosmids in *M. smegmatis*. In the goldfish model, *M. smegmatis* does not persist in tissue when inoculated at a dose of 10^7 organisms/fish. This is in contrast to *M. marinum,* which can be isolated from fish tissue throughout the course of a 56 day experiment. In this alternative approach, one can inject the fish with pools of the *M. marinum* cosmids in *M. smegmatis* and look for those which survive in the animal. A library of *M. marinum* cosmids in *M. smegmatis* can be obtained routinely, using standard, art-recognized procedures.

Once an insertionally mutated *M. marinum* bacterium has been identified as being a (putative) virulence mutant, a wild type *M. marinum* can be engineered to contain a more well-defined (*e.g.*, non-polar) mutation. The introduction of such a well-defined mutation into a new genetic background can confirm that the original phenotype was the result of the transposition event, rather than a secondary mutation. Furthermore, a well-defined mutation can be used to ascertain the presence, if any, of polarity effects. For example, the insertion of a transposon into a gene which is part of an operon can have polar effects on downstream genes in the operon. One method to determine if a given defect results from inactivation of the gene into which a transposon integrated, or if the actual virulence gene(s) lies downstream of the integration site, is to generate a small, in-frame, non-polar, deletion or insertion into a wild type correlate of the gene into which the transposon had

integrated. If such a mutant, when tested, for example as described above in the fish model, does not exhibit an avirulent phenotype, other genes in the operon can be mutated and analyzed in the same manner until one (or more) virulence genes are identified. That is, nucleic acid sequences which flank the integrated transposon can be cloned and sequenced in several sequential steps (*e.g.*, one can "walk" down an operon) until a virulence gene is identified. Of course, the invention includes genes which lie downstream of a gene in which a polar mutation results in an avirulent phenotype. Such genes can be considered to be "genes of the invention" or "genes identified by methods of the invention."

As a first step in performing site-specific mutagenesis of a gene of interest, it is preferable to isolate (*e.g.*, clone) at least a portion of the corresponding wild type gene. If the gene is part of an operon, some, if not all, of the other genes in the operon can also be isolated. As used in this application, the term "isolated" (referring, *e.g.*, to a gene or gene product, nucleic acid, protein, bacterium, etc.) means being in a non-naturally-occurring form. Methods to clone genes, particularly those containing a unique marker, are routine for one of ordinary skill in the art. (See, *e.g.*, Sambrook, J. *et al* (1989). *Molecular Cloning, a Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; Ausubel, F.M. *et al* (1995). *Current Protocols in Molecular Biology*, N.Y., John Wiley & Sons; Davis *et al.* (1986), *Basic Methods in Molecular Biology*, Elsevir Sciences Publishing, Inc., New York; Hames *et al.* (1985), *Nucleic Acid Hybridization*, IL Press; Dracopoli, N.C. *et al.*, *Current Protocols in Human Genetics*, John Wiley & Sons, Inc.; and Coligan, J.E. *et al.*, *Current Protocols in Protein Science*, John Wiley & Sons, Inc for many of the molecular biology techniques referred to in this application, including isolating, cloning, modifying, labeling, manipulating, sequencing, and otherwise treating or analyzing nucleic acid and/or protein.). In one method, clones comprising a gene(s) of interest can readily be identified and isolated from a wild type library (*e.g.*, a cosmid library, Bacterial Artificial Chromosome (BAC) library (Brosch, R. *et al* (1998). *Infect. Immun.* 66, 2221-2229; Philipp, W.J. *et al* (1996). *PNAS* 93, 3132-37), phage library, cDNA library, or the like), using conventional, routine, procedures in the art. Methods for subcloning a gene(s) of interest are also routine for one of ordinary skill in the art.

Example 6 describes a preferred embodiment of the invention, in which a hybridization probe corresponding to gene sequences flanking the site of transposon integration in an *M. marinum* mutant is used to screen a cosmid library of wild type *M. marinum* genes. Because many *M. marinum* genes are about 2 kb in size, and the average DNA insert in a cosmid library can be about 30-40 kb, it is likely that a cosmid clone so identified will contain the entire operon, if any, in which the gene of interest is located. It is understood, of course, that the genes and clones referred to in this application typically are double-stranded; therefore, a probe "corresponding to" a given sequence can be designed to hybridize to either of the strands of the DNA duplex, or to a nucleic acid (*e.g.*, RNA or cDNA) which is complementary to one strand of the duplex.

The term "a cloned gene," as used herein, can encompass not only the regions of DNA that code for a polypeptide but also regulatory regions of DNA such as regions of DNA that regulate transcription, translation and, for some microorganisms, splicing of RNA. Thus, a "gene" can include promoters, transcription terminators, ribosome-binding sequences and, for some organisms, introns and splice recognition sites. A cloned "gene" as used herein can be, *e.g.*, a genomic or a cDNA gene, or a rRNA or tRNA gene, or the like.

After a gene of interest, or a portion thereof, has been cloned, defined mutation(s) can be introduced into it, using methods of site-specific mutagenesis which are well-known in the art. Any type of mutation, for example those defined above, can be introduced into a cloned gene of interest. In a preferred embodiment, a wild type, cloned *M. marinum* virulence gene is mutated such that an insertion or deletion (ranging from about 3 bases to about 90% of the entire gene sequence, preferably about 99 to about 4000 bases, most preferably about 500 bases) is introduced in such a way that the coding sequences remain in phase (*i.e.*, the insertion or deletion is a multiple of 3 bases). In a most preferred embodiment, the mutation is an insertion of a nucleic acid fragment which comprises a kanomycin resistance marker. The site of the mutation can be chosen at will, but it is preferably in the 5'-terminal half of the gene. The availability of convenient restriction sites in the gene can simplify the introduction of mutations.

The mutated DNA can be reintroduced into the *M. marinum* genome by any of a variety of well-characterized methods. In a most preferred embodiment, the mutation is

introduced into the genome by allelic exchange (homologous recombination). Methods for using long linear recombination substrates for allelic exchange in Mycobacteria are provided, *e.g.*, in Balasubramanian, V. *et al* (1996). *J. Bacteriol.* 178, 273-279. Other methods for homologous recombination are found, *e.g.*, in Aldovini, A.R. *et al* (1993). *J. Bacteriol.* 175, 7282-7289; Norman, E. *et al* (1995). *Mol. Microbiol.* 16, 755-760; Baulard, A. *et al* (1996). *J. Bacteriol.* 178, 3091-3098; Marklund, B.I. *et al* (1995). *J. Bacteriol.* 177, 6100-6105; Ramakrishnan, L. et al (1997). *J. Bacteriol.* 179, 5862-5868; and U.S. Pat. No. 5,700,683.

Simultaneously with the characterization of a virulence defect in an *M. marinum* mutant, or prior or subsequent to such characterization, the gene which is disrupted by the transposon insertion can be identified and characterized. In one embodiment, regions flanking one or both sides of an integrated transposon are characterized by hybridization to a panel of selected sequences. In a most preferred embodiment, the flanking regions are sequenced in order to identify the gene which has been disrupted. Many sequencing methods are, of course, well-known to those of ordinary skill in the art. Example 5 describes two methods to sequence directly the flanking regions, as well as methods to first clone and then sequence such regions. In a most preferred embodiment, genomic sequences flanking a transposon are amplified using a strategy called ligation-mediated PCR (LMPCR) (Prod'hom *et al* (1998). *FEMS Microbiology Letters* 158, 75-81). Briefly, this method uses one primer specific for the known sequence (IS (insertion sequence) present on both ends of the transposon) and a second specific for a synthetic linker ligated to restricted genomic DNA. This method is illustrated in Figures 11 A and B. The size of the flanking regions which can be analyzed are limited by factors such as the fragment size that can be amplified by PCR, and can be readily determined by one of skill in the art. In a most preferred embodiment, a flanking region is about 100 to about 1,000 bases long.

The comparison of sequences of previously uncharacterized virulence genes in *M. marinum* to sequences in publicly available DNA and protein databases from a variety of sources (*e.g.*, GenBank, EMBL, DDBJ, SWISS-PROT, PRF, PDB, RefSeq, etc.) can aid in the identification of (functional) homologues, and can add insight into the role a virulence gene plays in the molecular pathogenesis pathways of mycobacteria in an animal host.

Optimal alignment of sequences may be conducted by the local homology algorithm of Smith and Waterman (1981). *Adv. Appl. Math.* 2, 482; by the homology alignment algorithm of Needleman and Wunsch (1970). *J. Mol. Biol.* 48, 443; by the search for similarity method of Pearson and Lipman (1988). *Proc. Natl. Acad. Sci.* 85, 2444; or by computerized implementations of these algorithms (*e.g.*, GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0., Genetics Computer Group, 575 Science Dr. Madison, Wis.) Other such computer programs include, *e.g.*, BLAST and FASTA (Altschul, S.F. *et al* (1990). *J. Mol. Biol.* 215, 403-410); BLASTX; TBLASTN; Gapped BLAST and PSI-BLAST (Altschul, S.F. *et al* (1997), *Nucleic Acids Res.* 25, 3389-3402). Alternatively, the sequences can be aligned by inspection. The best alignment (*i.e.*, resulting in the highest percentage of sequence similarity over the comparison window) generated by the various methods is selected. In a most preferred embodiment, the BLAST blastx program is used.

Typically, a polynucleotide sequence of interest is translated into all six possible reading frames and is searched with the NCBI Blast search, selecting blastx. This translated sequence is first run against the EMBL data base to identify functional homologs. Then, if desired, the sequence is searched with the advanced Blast program, against Mycobacterium sequences in particular. In a preferred embodiment, sequences identified by such a homology alignment exhibit substantial identity to the sequence of interest. Of course, any selected degree of sequence identity can be the basis of such a comparison, *e.g.*, about 30-50%, about 50-70% or about 70-90% sequence identity at the nucleotide or amino acid level.

The following terms are used to describe the sequence relationships between two or more polynucleotides or polypeptides: "reference sequence," "comparison window," "sequence identity," "percentage of sequence identity," and "substantial identity."

A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, a segment of a full-length cDNA or gene sequence given in a sequence listing, or may comprise a complete cDNA or gene sequence. Generally, a reference is at least about 10 nucleotides in length, frequently at least about 20 to 25 nucleotides in length, and often at least about 50 nucleotides in length. In a preferred embodiment, a reference sequence is

at least about 100 nucleotides in length, frequently at least about 150-300 nucleotides in length. Sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window," as used herein, refers to a segment of at least about 10 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least about 10 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions and deletions (*i.e.* gaps) of about 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences.

The term "sequence identity" means that two polynucleotide or polypeptide sequences are identical (*e.g.*, on a nucleotide-by-nucleotide or amino acid-by-amino acid basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I) or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "identical" in the context of two nucleic acid or polypeptide sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence.

The term "substantial identity" or "substantial similarity" indicates that a nucleic acid or polypeptide comprises a sequence that has at least about 90% sequence identity to a reference sequence, or preferably at least about 95%, or more preferably at least about 98% sequence identity to the reference sequence, over a comparison window of at least about 10 to about 100 or more nucleotides or amino acid residues. An indication that two polypeptide sequences are substantially identical is that one protein is immunologically reactive with antibodies raised against the second protein. An indication that two nucleic acid sequences are substantially identical is that the polypeptide which the first nucleic acids encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under selected high stringent conditions. High stringent conditions are sequence-dependent and will be different with different environmental parameters. Generally, high stringent conditions are selected to be about 5°C. to 20°C. lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, high stringent conditions will be those in which the salt concentration is at least about 0.2 molar at pH 7 and the temperature is at least about 60°C.

Analyses of the peptides or proteins which can be translated from flanking DNA sequences can be particularly informative for identifying functional homologues. The similarity between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide. Alignment procedures such as those discussed above can be used.

The sequencing and characterization of regions flanking thirteen transposons which have independently integrated into *M. marinum*, rendering the bacteria avirulent in the goldfish model, is shown in Example 9. At least six of the *M. Marinum* mutant genes are closely related to a previously identified functional homologue(s) from another organism, *e.g.*, a transcriptional regulator from *Streptomyces coelicolor* which belongs to the AraC family of transcriptional regulators; an integral membrane protein; polyketide synthase genes from *Streptomyces* and *Pseudomonas* bacteria; a sulfate adenylyltransferase with homology to diverse organisms including *Pyrococcus abyssi*, *Synechocytis*.sp., and *Bacillus subtilis;* a cysQ gene, or dhbF from *B. subtilis*. The possible significance of these functional properties for *M. marinum* virulence is discussed in Example 9.

The flanking sequences in *M. marinum* can also be compared in a similar manner to databanks of mycobacteria sequences, using the Advanced Blast search from NCBI and selecting Mycobacterium as the genome, and/or the complete sequence of *M. tuberculosis* (Cole, S.T. *et al* (1998). *Nature* 393, 537-558), in order to identify virulence genes in other mycobacteria. In a most preferred embodiment, this method can be used to identify virulence genes of *M. tuberculosis*. For example, Example 9 shows that the thirteen *M. marinum* virulence genes examined have functional homologues in *M. tuberculosis*.

Methods to clone such *M. tuberculosis* homologues are routine in the art. See, *e.g.*, Example 7.

Defined mutations can be introduced into cloned, putative virulence genes of *M. tuberculosis* genes by methods similar to those discussed above for mutagenizing cloned *M. marinum* genes. The mutations can be made in *M. tuberculosis* either before or after the corresponding mutations in *M. marinum* have been characterized. Any of the types of mutations described above can be introduced into an *M. tuberculosis* gene, including knockouts of a large portion, including the entire coding sequence, of the gene. In order to facilitate the generation of mutants in *M. tuberculosis*, conventional, routine procedures can be used to identify those regions of the *M. tuberculosis* gene which correspond to the site of mutation in the corresponding *M. marinum* gene. For example, corresponding active sites and/or functional domains can be identified by, *e.g.*, comparing the sequences or modeling the predicted protein structures. The mutated DNA can then be reintroduced into the *M. tuberculosis* genome by methods similar to those described above for reintroducing mutations into the *M. marinum* genome. Several such methods are described in Example 7. In a most preferred embodiment, the defined mutation is reintroduced into the *M. tuberculosis* genome by homologous recombination using a long linear recombination substrate. The phenotypic effect of an *M. tuberculosis* mutation can be determined routinely with one of several available animal models for this organism, including, *e.g.*, the infection models with guinea pig (Collins, D.M. *et al* (1995). *PNAS* 92, 8036-8040; B. Bloom, ed., (1994). *Tuberculosis: Pathogenesis, Protection, and Control*, ASM Press, Washington, D.C. Chapter 9); mouse and rabbit (B. Bloom, ed., *ibid*, Chapters 8 and 10, respectively); and monkey (Walsh et al (1996). *Nature Medicine* 2, 430-436).

The invention encompasses virulence genes *(e.g.*, isolated virulence genes) as described elsewhere herein, from *M. marinum* and/or *M. tuberculosis*, which are identified by the methods of the invention, and/or variants (*e.g.*, naturally- or non-naturally-occurring modifications, mutations, polymorphisms, etc.) or fragments thereof. By a "variant" of a gene or fragment is meant, as used herein, a replacement, deletion, insertion or other modification of the gene or fragment. It is preferred that the variant has at least about 70% sequence identity, more preferably at least about 85% sequence identity, most preferably

at least about 95% or 98% sequence identity with the gene or fragment. The degree of similarity can be determined using any of the methods disclosed herein. By a "fragment" of a gene is meant a single strand or double stranded nucleic acid (*e.g.*, oligonucleotide) of a size smaller than that of the gene, obtained by any of a variety of conventional means, *e.g.*, digestion with restriction enzymes, PCR amplification, synthesis with an oligonucleotide synthesizer, synthesis with a DNA or RNA polymerase, or the like. Such fragments can be used, for example, to diagnose the presence of a gene in a sample of interest, *e.g.*, by serving as a hybridization probe or a PCR primer. Such diagnostic assays can be set up and performed by routine, conventional procedures in the art. In another embodiment, such fragments can be used to screen for virulent strains of bacteria, *e.g.*, bacteria which comprise a polynucleotide that encodes a particular virulence gene or a fragment thereof. Of course, full-length virulence genes of the invention and variants thereof can also be used in diagnostic assays.

The invention also encompasses polynucleotides which are complementary to a gene of the invention or fragment thereof, or which hybridize to such a gene or fragment under selected (*e.g.*, high) stringency conditions. For example, the invention encompasses an oligonucleotide complementary to a portion of a virulence gene which can be used, *e.g.*, as an antisense oligonucleotide to regulate expression of the gene, *e.g.*, in a method of therapy. Methods to make and use antisense molecules of this type are conventional and routine, and are presented, *e.g.*, in U.S. Pat. Nos. 5,876,931 and 5,585,479 and in references cited therein. Similarly, ribozymes comprising such fragments can be used in a method of treatment. Methods of making and using ribozymes are also conventional in the art.

Of course, the genes and fragments discussed herein can be any form of polynucleotide or nucleic acid, *e.g.*, naturally occurring, synthetic or intentionally manipulated polynucleotides, wherein nucleotide bases or modified bases are linked by various known linkages, *e.g.*, ester, phosphodiester, sulfamate, sulfamide, phosphorothionate, phosphoroamidate, methyl phosphonate, carbamate, or other bonds, depending on the desired purpose, *e.g.*, resistance to nucleases, such as RNAse H, improved *in vivo* stability, etc. Various modifications can be made to nucleic acids, such as attaching detectable markers (*e.g.*, avidin, biotin, radioactive or fluorescent elements,

ligands), or moieties which improve hybridization, detection or stability. The polynucleotides can be DNA, cDNA, RNA, PNA, synthetic nucleic acid, modified nucleic acid, or mixtures thereof. Polynucleotides can be of any size, *e.g.*, ranging from short oligonucleotides to large gene clusters or operons. Either or both strands of a double strand nucleic acid are included.

The invention also encompasses peptides or polypeptides encoded by and/or expressed from *M. marinum* and/or *M. tuberculosis* genes identified by the methods of the invention, and/or variants or fragments thereof, and products which are generated by such peptides or polypeptides. The term "genes identified by the methods of the invention" encompasses any gene in a given operon, a mutation in one of whose genes results in an avirulent phenotype (*e.g.*, the gene can be a downstream gene whose expression is diminished or abolished because of an upstream polar mutation, or a gene whose gene product interacts with another gene product of the operon, etc.).

The peptides or polypeptides can be isolated (*e.g.*, purified) from bacteria directly, or they can be expressed recombinantly and isolated (*e.g.*, purified) from recombinant organisms. Methods of isolating, purifying and sequencing naturally produced or recombinantly produced peptides and polypeptides are conventional and routine in the art. The genes can be cloned into any of a variety of expression vectors. The sequences to be expressed can be genomic sequences, *e.g.*, subcloned sequences from a cosmid library as described in Example 6, or they can be corresponding cDNA sequences, obtained by conventional means. In some cases, it may be desirable to express a fragment of a gene, or more than one gene, *e.g.*, as many as the genes of an entire operon. Vectors and appropriate regulatory elements for expressing genes in a variety of cell types or hosts, including prokaryotes, yeast, and mammalian, insect and plant cells, and methods of cloning and expressing genes or gene fragments, are routine in the art and are discussed, *e.g.*, in U.S. Pat. Nos. 5,876,931, 5,700,683, 4,440,859, 4,530,901, 4,582,800, 4,677,063, 4,678,751, 4,704,362, 4,710,463, 4,757,006, 4,766,075 and 4,810,648.

The invention also encompasses a host transformed to express a peptide or polypeptide of the invention, or a host which is mutated so the expression of a peptide or polypeptide of the invention is disrupted (*e.g.*, inhibited), or progeny of such hosts.

"Variants" of the peptides or polypeptides are also included in the invention, *e.g.*, insertions, deletions and substitutions, either conservative or non-conservative, where such changes do not substantially alter the normal function of the protein. By "conservative substitutions" is meant by combinations such as Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr. Variants can include, *e.g.*, homologs, muteins and mimetics. Many types of protein modifications, including post-translational modifications, are included. See, *e.g.*, modifications disclosed in U.S. Pat. No. 5,935,835.

"Fragments" of the peptides or polypeptides are also included in the invention. These fragments can be of any length. In a preferred embodiment, a fragment is functional (*e.g.*, has biological activity, can inhibit or enhance the activity of a protein or other substance, contains one or more immunogenic epitopes, etc.). In a most preferred embodiment, the fragment contains all or a subset of the amino acids of SEQ ID NOs: 5, 7, 9, 12, 15, 20, 22, 24, 26, 28, 30, 32-34, 36-38, 40 or 42.

Among the polypeptides of particular interest are polyketide synthases. Example 9, for example, shows that an *M. marinum* virulence gene identified by the method of the invention, and an *M. tuberculosis* homologue of it, appear to be polyketide synthase genes. As is well-known, many polyketides have therapeutic value (for human, veterinary, or aquaculture uses). For example, polyketides have been shown to function as antibiotics, chemotherapeutic agents or immunosuppressive agents, *e.g.*, in transplant patients. The invention includes the generation and/or isolation (*e.g.*, purification) of polyketide synthases encoded by virulence genes identified by the method of the invention, as well as polyketides produced by those synthases. The polyketides can be generated by recombinant means, isolated from non-recombinant bacteria, or produced synthetically. Methods for making, isolating and purifying polyketides are routine and well-known in the art.

Recombinantly expressed polypeptides of the invention can also be used to confirm that a particular virulence gene is responsible, at least in part, for a pathogenic phenotype in an organism - that is, to confirm Koch's postulates. Example 8 shows how a recombinantly expressed *M. marinum* putative virulence gene can be used to complement a mutant bacterium which is defective in that gene, and to restore a virulent phenotype in fish infected by the complemented mutant.

Virulence genes of the invention and peptides thereof can contain antigenic epitopes. The invention also encompasses antibodies, including polyclonal or monoclonal antibodies, or fragments of polyclonal or monoclonal antibodies, which are generated in response to such epitopes. Such antibodies can be used, *e.g.*, in diagnostic assays to detect the presence of a mycobacterium, to identify virulent strains of bacteria, or in methods to treat disease conditions caused or exacerbated by a virulence protein (*e.g.*, passive immunization), following routine, art-recognized procedures.

The invention also encompasses an avirulent mycobacterium, preferably *M. marinum* and/or *M. tuberculosis*, which harbors one or more mutation(s) in one or more virulence gene(s) identified by the methods of the invention, or a pharmaceutical composition which comprises such a bacterium and a pharmaceutically acceptable carrier.

In a preferred embodiment, the avirulent bacterium is introduced into a host (*e.g.*, a fish, cow or human) in order to elicit an immune response. Because the bacterium is avirulent (*e.g.*, attenuated), it is expected to be suitable for administration to a host in need of treatment, but it is also expected to be antigenic and to give rise to an immune response, preferably a protective immune response. For such a use, it is preferred that the mutation is substantially non-revertable, *e.g.*, a deletion or frame-shift mutation. To ensure non-revertability, it is preferable that a bacterium comprises at least two or three such mutations, preferably in different genes. A small deletion mutant would be expected to provide antigenic epitopes in the portion of the protein which lies downstream of the deletion, even though the protein, itself, is not functional with respect to virulence.

Another embodiment of the invention is a vaccine comprising a suitable avirulent mycobacterium of the invention and a pharmaceutically acceptable carrier. By vaccine is meant an agent used to stimulate the immune system of a living-organism so that protection against future harm is provided. Immunization refers to the process of inducing an antibody and/or cellular immune response in which T-lymphocytes can either kill the pathogen and/or activate other cells (*e.g.*, phagocytes) to do so in an organism, which is directed against a pathogen or antigen to which the organism has been previously exposed. The term "immune response," as used herein, encompasses, for example, mechanisms by which a multi-cellular organism produces antibodies against an antigenic material which invades the cells of the organism or the extra-cellular fluid of the organism. The antibody

so produced may belong to any of the immunological classes, such as immunoglobulins A,D,E,G or M. Other types of responses, for example cellular and humoral immunity, are also included. Immune response to antigens is well studied and widely reported. A survey of immunology is given *e.g.*, in Roitt I., (1994). *Essential Immunology*, Blackwell Scientific Publications, London. Methods in immunology are routine and conventional (see, *e.g.*, in *Current Protocols in Immunology*; Edited by John E. Coligan *et al.*, John Wiley & Sons, Inc.).

Methods of formulating, testing, optimizing and administering vaccines of the invention are routine and conventional, and are described, *e.g.*, in U.S. Pat. Nos. 5,876,931, 5.700,683, and references cited therein, and in "New Generation Vaccines, edited by M.M. Levine *et al,* 2nd edition, Marcel Dekker, Inc., New York, NY, 1997." Active immunization of a patient (*e.g.*, human, fish, cow, etc.) is preferred. In this approach, one or more mutant bacteria are prepared in an immunogenic formulation containing suitable adjuvants and carriers and administered to the patient in known ways. Suitable adjuvants include Freund's complete or incomplete adjuvant, muramyl dipeptide, the "Iscoms" of EP 109 942, EP 180 564 and EP 231 039, aluminum hydroxide, saponin, DEAE-dextran, neutral oils (such as miglyol), vegetable oils (such as arachis oil), liposomes, Pluronic polyols or the Ribi adjuvant system (see, for example GB-A-2 189 141). "Pluronic" is a Registered Trade Mark. The patient to be immunized is a patient requiring to be protected from the disease caused by, or exacerbated by, the virulent form of the bacterium.

The aforementioned avirulent bacteria of the invention or a formulation thereof may be administered by any conventional method including oral and parenteral (*e.g.*, subcutaneous or intramuscular) injection. The treatment may consist of a single dose or a plurality of doses over a period of time. While it is possible for an avirulent bacterium of the invention to be administered alone, it is preferable to present it as a pharamaceutical formulation, together with one or more acceptable carriers. The carrier(s) must be "acceptable" in the sense of being compatible with the avirulent microorganism of the invention and not deleterious to the recipients thereof. Typically, the carriers will be water or saline which will be sterile and pyrogen free.

It will be appreciated that a vaccine of the invention, depending on its bacterial component, may be useful in the fields of human medicine, veterinary medicine, or aquaculture. A vaccine for fish against *Mycobacterium marinum* could be of particularly significant economic importance. *Mycobacterium marinum* causes tuberculosis in more than 150 species of both salt-water and fresh-water fish, among them salmonid trout (*salmo gairdneri, salmo trutta, oncorhynchos mykiss*), striped bass, tilapia, etc. Aquaculture facilities infected with *M. marinum* suffer from a constant mortality rate over a long period of time accompanied by severe economic losses, which could be ameliorated with such a vaccine. A vaccine against *M. tuberculosis* could, of course, be a significant weapon in the battle against tuberculosis, which is wide-spread in human populations.

Vaccines encompassed by the invention also include killed bacterial vaccines; subunit vaccines comprising a virulence protein(s) of the invention (*e.g.*, a wild type or mutant protein(s), or a variant(s) thereof), or an antigenic fragment(s) thereof; bacteria which produce or are capable of producing such virulence proteins or fragments; and DNA vaccines comprising a nucleic acid which encodes such a virulence protein or fragment thereof. Methods of making and using such vaccines are routine and conventional in the art. For methods of making and using DNA vaccines, see, *e.g.*, U.S. Pat. No. 5,589,466.

An avirulent bacterium of the invention can also be used as a "carrier" for the expression of one or more cloned heterologous gene(s) or fragments thereof. For example, an avirulent *M. marinum* organism can be used to express a secreted or surface-expressed heterologous peptide or polypeptide in fish, and an avirulent *M. tuberculosis* organism can be so used in humans. The avirulent bacterium can be used to express, *e.g.*, an allergen, or an antigenic epitope from another pathogen, for which the modified bacterium can act as a vaccine. In a preferred embodiment, the heterologous gene is inserted at or near the position at which the transposon was inserted in an avirulent mutant, or at or near the site of the more "well-defined" avirulent mutation. Methods to clone heterologous genes are routine, as are methods to express them in a host. Methods of making and using such carriers are disclosed, *e.g.*, in U.S. Pat. Nos. 5,876,931 and 5,424,065.

The invention also encompasses a method for identifying an agent which reduces the ability of a microorganism to survive in a host, *e.g.*, an anti-mycobacterial agent which inhibits expression of a virulence gene, or which attacks products produced directly or

indirectly by a virulence gene. In a preferred embodiment, such an agent can be used to treat a disease caused by, or exacerbated by, a virulence gene of the invention. One such method, as disclosed, *e.g.*, in U.S. Pat. No. 5,876,931, is to generate a bacterium which over-expresses the virulence gene, and then to identify an agent which reduces the viability or growth of a wild type cell but not the cell overexpressing the gene, in a host. Methods to generate the over-expressing strain, and to perform such screening procedures, are routine and are described, *e.g.*, in U.S. Pat. No. 5,876,931. Other methods to screen for anti-mycobacterial drugs are routine and are described, *e.g.*, in U.S. Pat. No. 5,700,683.

The invention also relates to a method of screening vaccine candidates for human tuberculosis in the fish model. In one embodiment, based on the assumption that *M. marinum* bacteria may be suitable for human vaccines, goldfish can be inoculated with an *M. marinum* vaccine candidate of interest. The fish are then challenged with fully virulent *M. marinum* at a dose capable of establishing disease. A vaccine which, when inoculated into a fish, protects the fish from subsequent virulent challenge by the fish failing to develop disease symptoms is a candidate for a human vaccine. In another embodiment, a putative virulence gene of *M. tuberculosis* is selected, and a mutation is made in the *M. marinum* homologue of that gene. The mutant *M. marinum* is then tested as a vaccine candidate, using the goldfish model as above.

Brief Description of the Figures

Fig. 1 shows the median survival time (MST) of fish inoculated with *M. marinum*. The median survival time of fish (days) inoculated with *M. marinum* at doses indicated per fish is compared to a phosphate buffered saline (PBS) control. *survival to endpoint of experiment, 56 days.

Fig. 2 shows a comparison of the growth of *M. marinum* in liver, spleen and kidney. The inoculum is 10^7 CFU/fish. Results are given as geometric means ± standard error for eight fish per time point.

Fig. 3 shows a comparison of mean cumulative granuloma scores (MCGs) over time of fish infected with 10^7 CFU of *M. marinum* organisms. The results are given as a vertical box plot, with horizontal lines marking the median 10^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentile points of GSs for eight animals at each time point. The mean of each group is

represented by a thick line. At 2 weeks, the median 50th percentile and mean values are the same.

Fig. 4 shows a survival curve of goldfish inoculated with 10^8 CFU of *M. marinum* 1218R (wild type) or 1218S (mutant).

Fig. 5 shows the modification of pYUB285 with transposon tags. Bg is BglII; Bam is BamH1; H is HindIII; IR are inverted repeats which mark the boundaries of the transposon; ORFR and ORFA are transposon genes; *aph* is the gene for kanomycin resistance; *ori*E is the *E. coli ori*; and Δ*ori*M is the disabled mycobacterial *ori*.

Fig. 6 shows the construction of an *M. marinum* signature-tagged mutant library.

Fig. 7 shows a schematic diagram of an *M. marinum* mutant library screen in the goldfish model.

Fig. 8 shows a survival curve of *M. marinum* mutant 41.2.

Fig. 9 shows a survival curve of *M. marinum* mutant 80.1.

Fig. 10 shows a survival curve of *M. marinum* mutant 86.1.

Figs. 11A and B illustrate ligation-mediated PCR.

Fig. 12 shows Competitive Indices of *M. marinum* mutants 32.2, 60.2, 62.2, 67.1, 80.1, 86.1, 42.2, 80.8 and 68.6.

Fig. 13 shows a survival curve of *M. marinum* mutant 67.1.

Fig. 14 shows a survival curve of *M. marinum* mutant 39.2.

Fig. 15 shows a survival curve of *M. marinum* mutant 42.2.

Examples

Example 1. Properties of the *M. marinum*/goldfish model

A. Median Survival Time and LD_{50}.

To determine the median survival time of goldfish after inoculation with *M. marinum* strain ATCC 927, groups of 20 to 32 fish were inoculated intraperitoneally with 10^9, 10^8, or 10^7 colony forming units (CFU). The median survival time of goldfish inoculated with *M. marinum* was dose dependent, with survival time decreasing with increasing doses of bacteria. The median survival time of fish was 4, 10, and >56 days (the endpoint of the experiment) with inocula of 10^9, 10^8, or 10^7 *M. marinum* organisms, respectively. All fish inoculated with 10^7 CFU or less survived to the end point of the experiment (56 days). The control fish group, inoculated with PBS in 5 separate experiments, had a total of two premature deaths, one at 8 and one at 19 days post-inoculation, from a total of 55 fish. The remainder of the control fish survived to 56 days, the endpoint of the experiment (See Figure 1). The LD_{50} at 1 week postinfection with *M. marinum* was 4.5×10^8 (calculated by the method of Reed & Muench, 1938. *Am. J. Hyg.* 27, 493-497).

B. Mycobacterial recovery from fish organs.

To assess the ability of *M. marinum* to persist in goldfish tissue, the liver, spleen, and kidneys from each sacrificed fish were collected for bacteriological examination. *M. marinum* was recovered from all organs of fish in the 10^9 or 10^8 CFU inoculum groups. In fish inoculated with 10^7 CFU, *M. marinum* was recovered from 96% of the examined organs.

The fate over an 8 week period of the *M. marinum* ATCC 927 strain in the livers, spleens, and kidneys of fish inoculated with 10^7 CFU was followed. (See Figure 2). There was a significant positive linear relationship between time postinoculation and colony recovery in the liver ($P < 0.001$); for the spleen and kidneys, the relationship was positive but did not reach statistical significance ($P = 0.054$ and $P = 0.091$, respectively). Between 8 and 16 weeks postinoculation, *M. marinum* persisted in the tissue with no significant change in the colony counts. In addition, in the 10^2 to 10^6 CFU inoculum groups, *M marinum* was isolated from at least one organ from all infected fish.

C. An acute and chronic form of mycobacterial infection.

The pathology of infected fish was dependent on the inoculum dose and the time postinfection of animal sacrifice. Fish infected with either 10^9 or 10^8 CFU of *M. marinum* organisms suffered from anorexia, sluggish movement, and loss of equilibrium.

The histopathology of fish infected with 10^9 and 10^8 CFU was characterized by severe peritonitis and necrosis as compared to control fish. The peritoneum was filled with inflammatory cells consisting of lymphocytes, macrophages, fibrous connective cells as well as with degenerating cells and bacteria. The mean cumulative granuloma score (MCGS) for these 2 groups was similar (0.2 for the 10^9 CFU group and 0.9 for the 10^8 CFU group). In the 10^8 CFU inoculum group, granuloma formation was more likely to be found in animals which survived more than 2 weeks postinoculation.

When examined at 2 weeks, 6 of 8 fish in the 10^7 CFU group had moderate to severe peritonitis. Unlike the 10^8 and 10^9 CFU inoculum groups which succumbed to infection, the 10^7 CFU inoculum group survived the infection, and by 4 to 6 weeks postinoculation, the acute peritoneal inflammation was replaced by a chronic inflammatory state. Fish inoculated with 10^7 CFU demonstrated granuloma formation in all organs evaluated (MCGS of 5.0), including the peritoneum and pancreas, liver (*e.g.*, onion ring granuloma composed of epithelioid macrophages surrounding a necrotic center), spleen, trunk kidney, head kidney, heart and intestine. Pleomorphic granulomas (necrotizing, non-necrotizing and caseous) were seen. The necrotizing granulomas were characterized by a central area of necrosis surrounded by macrophages, epithelioid cells, and thin fibrous connective tissue. Frequently, caseous necrosis was present in the central area of the granuloma. Granulomas containing foamy macrophages were also seen. Occasionally, Langhans and foreign body type giant cells were observed. In addition, acid fast bacilli could be demonstrated with the modified Ziehl-Neelsen stain. Melanomacrophage centers were seen in a few cases.

The chronic inflammatory response of fish towards *M. marinum* was time dependent, as seen by the increment in mean cumulative granuloma scores (MCGSs) with time in animals inoculated with 10^7 CFU (See Figure 3) up to 8 weeks. From 8 to 16

weeks postinoculation, there was no significant change in MCGSs (5.0 and 5.7 respectively).

D. Minimum infectious dose (MID).

To estimate the lowest possible dose of *M. marinum* able to establish infection in goldfish, groups of four fish were inoculated with *M. marinum* ATCC 927 at doses of 10^6, 10^5, 10^4, and 10^2 CFU. Granuloma formation was seen in 25% of the goldfish by 4 weeks and in 88% by 8 weeks postinfection with a dose of 6.3×10^2 CFU or higher (Table 1). The minimum number of organisms required to establish infection in goldfish appears to be approximately 600 CFU.

Table 1. MID of *M. marinum* ATCC 927

Inoculum (CFU/fish)	No. positive[a]		MCGS
	4Wk	8Wk	
1.2×10^6	1 / 2	1 / 2	5.0
3.0×10^5	0 / 2	2 / 2	5.5
2.4×10^4	1 / 2	2 / 2	1.5
6.3×10^2	0 / 2	2 / 2	4.5

[a]Number of granuloma-positive animals per total number of animals at 4 and 8 weeks postinoculation.

Mycobacterial virulence assay.

The relative virulence of different strains of *M. marinum*, isolated from both human and animal origin, was assessed. Three mycobacterial strains, *M. marinum* ATCC 927, M and F-110, were inoculated into goldfish at 10^8 CFU. The median survival times of *M. marinum* M, ATCC 927, and F-110 were similar, ranging from 4 to 10 days.

Example 2 - Differentiation of an avirulent *M. marinum* mutant from the wild type in the goldfish model

The goldfish model can differentiate between virulent and avirulent *M. marinum* organisms. A comparison of such a pair of strains is shown in Figure 4. The *M. marinum* strains designated 1218R (wild type, aka ATCC 927) and 1218S (avirulent mutant) were inoculated into groups of 5 to 9 goldfish in two separate experiments at an inoculum dose of 1.4 to 4 x 10^8 CFU. The median survival time of goldfish inoculated with *M. marinum* 1218R organisms was 3 days compared to 28 days (endpoint of experiment) with *M. marinum* 1218S organisms (See Figure 4). The mutant 1218S also failed to persist in the mouse macrophage model. This experiment shows that the fish mycobacteriosis model can allow the identification of *M. marinum* virulence genes.

Example 3 - Signature-tagged mutagenesis, and the generation of a library

A. Construction of a master bank of signature-tagged transposons

As an initial step in creating a bank of signature-tagged transposons, plasmid pAT30 is generated (see Figure 5). A unique restriction site (*Bgl*II) is introduced into the mycobacterial transposon delivery vector pYUB285 between ORFA and *aph*. The vector is a suicide vector in mycobacteria because of inactivation of the mycobacterial origin of replication by an internal deletion. A kanamycin resistance gene (*aph*) inserted into IS*1096* allows for a library of insertions in the mycobacterial genome to be generated upon electroporation of the plasmid followed by selection for kanamycin.

To generate a collection of signature tagged transposons to be inserted into pAT30, primers P5 (5'-CTAGGTACCTACAACCTC-3') (SEQ ID NO: 1) and P3 (5'-CATGGTACCCATTCTAAC-3') (SEQ ID NO: 2) and the template RT1 oligonucleotide (5'-CTAGGTACCTACAACCTCAAGCTT-$[NK]_{20}$-AAGCTTGGTTAGAATGGGTACCATG-3') (SEQ ID NO: 3) are prepared by conventional, routine methods, preferably using a commercially available oligonucleotide synthesizer. The 5' ends of primers P5 and P3 have *Bam*HI sites. The template RT1 oligonucleotide is similar to that designed by Hensel *et al.*, with a variable central region

$(NK)_{20}$ flanked by arms of invariant sequences. The invariant arms allow the sequence tags to be amplified in a PCR with the use of primers P3 and P5. The variable region is designed to ensure that the same sequence occurs only about once in 2×10^{17} molecules. PCR is performed, using standard, routine methods (see, *e.g.*, Innis, M.A. *et al.*, eds. *PCR Protocols: a guide to methods and applications*, 1990, Academic Press, San Diego, CA) to generate and amplify double stranded, 90 bp signature tags. The PCR amplified tags are digested with *Bam*HI, gel purified, and then ligated to the *Bgl*II digested, dephosphorylated (calf intestinal phosphatase, New England BioLabs, Inc.) pAT30 plasmid. *E. coli* DH5α is transformed with this ligation mixture and plasmids from 800 individual clones are isolated, arrayed in 96 well microtiter plates, and transferred to nylon membranes. These plasmids are analyzed for hybridization and tag amplification efficiency. In this example, ninety-six plasmids that are hybridization and amplification efficient are chosen for the master plasmid collection. The master plasmids are screened for cross hybridization with other plasmids in the master plasmid collection and any cross-hybridizing plasmids are eliminated until the collection has no cross hybridizing members. Of course, a master plasmid collection of any size can be constructed by this method. Methods for carrying out STM mutagenesis and isolating bacterial virulence mutants are described, *e.g.*, in Hensel *et al* (1995). *Science* 269, 400-403 and U.S. Pat. No. 5,876,931.

B. Optimization and initial characterization of *M. marinum* transposition

Several protocols for the preparation of competent cells from *M. marinum* are evaluated. The strains tested are ATCC 927 (fish isolate) and *M. marinum* strain M (human isolate). Electrocompetent cells are prepared from *M. marinum* cells grown to different growth phases at different temperatures in the presence of ethionamide or cycloheximide. Mycobacterial cells are transformed by electroporation with the replicative *Escherichia coli*- mycobacteria shuttle vector, pYUB18 (Jacobs, W.R. *et al* (1991). *Methods Enzymol* 204, 537-555), as well as the suicide vectors pYUB285 (McAdam R.A. *et al* (1995). *Infect. Immun.* 63, 1004-1012) and pUS252, carrying the transposable elements, IS*1096* and IS*6110*, respectively (Dale, J.W. (1995). *Eur. Respir. J.* 8, 633s-648s). Mutants of *M. marinum* are recovered on 7H10 agar plates supplemented with kanamycin. Transformation and transposition efficiencies under different protocols are compared, using routine, art-recognized procedures. See, *e.g.*, McAdam *et al* (1995).

Infec. Immun. 63, 1004-1012 and Cirillo, J. D. *et al* (1991). *J. Bacteriol.* 173, 7772-7780. Southern hybridization analysis is performed on mycobacterial mutants to confirm the transposition events. These analyses show that: 1) competent cells prepared at room temperature from late-exponential growth phase organisms yield a higher transposition efficiency than cells prepared at 4°C or from early-or mid-exponential growth phase organisms; 2) the highest efficiency for transposition is 10^2-10^3 cfu per μg of plasmid DNA; and 3) the IS*1096*-derived transposon is best able to efficiently mutagenize *M. marinum*.

To confirm that *M. marinum*-kanamycin resistant colonies are not spontaneous mutants, colonies recovered after electroporation with the non-integrating, replicative vector, pYUB18, are analyzed; the plasmid pYUB18 is successfully isolated from 6 separate transformants and is identified by restriction enzyme mapping. This indicates that the transformants are not spontaneous mutants. In another experiment, 35 randomly selected mutants recovered from electroporation of the suicide vector, pYUB285 are examined by Southern analysis to determine whether transposition is random in the *M. marinum* chromosome. All tested transposon mutants yield a single band, located in a different position on the Southern blot, consistent with random integration of a single copy of IS*1096* into the *M. marinum* genome. Evaluation of 10 mutants obtained in a single electroporation experiment shows that each mutant is inserted into a different part of the *M. marinum* genome, indicating that the mutants from a given electroporation do not represent siblings.

C. Generation of an *M. marinum* mutant library

An *M. marinum* mutant library is generated by electroporating individual members of the 96 master plasmid collection into *M. marinum* bacteria (See Figure 6). *M. marinum* electrocompetent cells are prepared from a 100 ml culture grown to late exponential phase ($O.D._{600}$ =1.6 to 1.8). Bacteria are washed three times at room temperature with 10% glycerol and then suspended in 1 ml 10% glycerol and distributed to 0.2 cm gap electroporation cuvettes (Bio-Rad Laboratories). Electroporation is performed at room temperature using a Gene Pulser (Bio-Rad Laboratories) with parameters of 2.5 kV, 25 μF, and 800 Ω. Electroporated cells are rescued by growth overnight in 7H9 broth with 10% albumin-dextrose complex enrichment (ADC) (52) at 30°C and plated on 7H10 agar with

kanamycin (20μg/ml) and incubated at 30°C. Mutants appear 1 to 2 weeks after plating. Mutants from each electroporation are named for the master plasmid used for transposon delivery (pAT30-1 plasmid yields mutants 1.1, 1.2, etc.). In this example, 960 mutants are isolated, 10 mutants per master plasmid. Of course, more mutants can be isolated per each master plasmid, and the 96 (or additional) master plasmids can be used to generate additional mutants.

Example 4 - Screening an *M. marinum* library for potential avirulent mutants, using the goldfish model

A. Screening for mutants which show reduced viability in the goldfish host

The *M. marinum* library obtained in Example 3 is screened for mutants which exhibit a reduced ability to survive in the goldfish model. The library of *M. marinum* transposon-tagged mutants is screened in pools; in this example, each pool has 48 mutants (See Figure 7). Each of the mutants in a given pool is marked with a unique DNA tag (*i.e.* they are derived from 48 of the 96 master plasmids). To generate an input pool, mutants that make up the pool are grown in individual wells of a 96-well microtiter plate containing 7H9 broth with ADC and kanamycin (20μg/ml) at 30°C until they reach $O.D._{600}$= 0.6-0.8. The mutants are then pooled and an aliquot is removed for amplification using colony PCR (input pool probe). The remaining pooled bacterial cells are centrifuged, resuspended in phosphate buffered saline (PBS) to an inoculum dose of about $2x10^7$ cfu/ml, sonicated for 3 minutes, and injected into three fish. The fish are sacrificed at 7 days postinoculation and spleen, liver and kidney are harvested. The mutants that have reached and multiplied within these organs are recovered by plating homogenates of the organs onto laboratory medium. The recovered mutants from a given organ are combined and an aliquot is used for amplification using colony PCR (output pool probe). The products of the input and output pool amplification are used in a second PCR amplification using α-^{32}P dCTP to generate two radiolabeled probes. The amplified probes consist of a central variable region (the unique DNA tag) flanked by arms of invariable sequences which permit amplification of any tag using a defined set of primers. The arms are released by digestion with *Hind* III and the radiolabeled tags are used to probe replicate

membranes from the master plasmid collection. Because of the complex structure of the mycobacterial cell wall and difficulties encountered in mycobacterial colony hybridization, in this example the amplified tags are used as probes to a dot blot containing the master plasmid collection. Hybridization to other forms of the master plasmid collection can, of course, be used. Tags from mutants that hybridize to the probe from the input pool (Figure 7, membrane 1) but not to the probe from the output pool (Figure 7, membrane 2) represent mutants which are unable to survive or compete in the fish model. Such mutants are designated as potential virulence mutants.

The pools of mutants recovered from different organs are kept separate, in order to characterize virulence mutants with regard to the organs examined. In some cases, mutations necessary for survival at different points in the pathogenesis of this organism can be identified, since the mechanisms necessary for survival in liver, spleen and kidney, or in other organs, may differ. The pools of mutants recovered form different fish are also kept separate. Mutants from two fish are used independently to produce an output pool probe and are independently hybridized to replica membranes to confirm reproducible identification of potential virulence mutants from a given experiment.

B. Confirming that the mutants are avirulent by examining individual mutants in the goldfish model.

M. marinum transposon mutants that reproducibly hybridize to the input pool probe but not to the output pool probe are examined individually in the goldfish model. An inoculum dose of 10^8 bacteria in 0.5 ml per fish is used to inoculate 3 fish per mutant. A control group of fish is simultaneously inoculated with *M. marinum* ATCC 927 (wild type) at the same dose as the mutants and with PBS as a negative control. The median survival time (MST) of goldfish inoculated with the wild type at this dose is 10 days. If the MST for a given mutant is greater than that of the wild type, this confirms that the mutant may have the transposon inserted into a virulence gene. When a mutant-inoculated fish survives for 35 days, it is sacrificed and examined for histopathology; and portions of the liver, spleen and kidney are homogenized and plated for colony counts. These mutants are then inoculated into fish to determine the LD_{50}. Three fish per mutant per dose are injected with 10^8, 5×10^7, or 10^7 CFU bacteria. The LD_{50} for each mutant is evaluated at 1 week postinoculation and calculated by the method of Reed and Meunch (1938. *Am. J. Hyg.* 27,

493-497). The LD_{50} at 1 week for the wild type strain is 4.5×10^8 CFU bacteria per fish. The LD_{50}, Competitive Index, and/or pathology for each mutant is compared to that of the wild type strain.

Competitive index: The competitive index may be used as a measure of the attenuation of a mutant with respect to a wild type strain. Mutant and wild type strains are mixed together in the inoculum. Animals are inoculated with the mixture and 2 weeks post-inoculation the animals are sacrificed. The liver of the animal is removed, homogenized, and the colony counts in the tissue are determined for both the mutant and wild type strains. The two strains are distinguished because the mutant is kanamycin resistant while the wild type is kanamycin sensitive. Mathematically, the competitive index is defined as the output ratio of mutant to wild type bacteria, divided by the input ratio of mutant to wild type bacteria. A mutant which has full virulence with respect to the wild type should not be out competed by the wild type and the competitive index should be 1.0.

Histopathology examinations: Portions of the liver, spleen and kidney along with peritoneum, heart, pancreas, or other organs evident to one of skill in the art, are fixed in 10% neutral buffered formalin for routine embedding in paraffin. Five μm thin sections of the paraffin fixed tissues are prepared with a rotary microtome (American Optical, Buffalo, NY). After dewaxing, the sections are stained for acid fast bacilli with modified basic fuchsin stain and counterstained with methylene blue or stained with hematoxylin and eosin.

Colony counts in organ homogenates or the ability to induce granuloma formation: These parameters can identify virulence defects which are more subtle than one which causes the MST to change. Mutants identified in the screening protocol as failing to survive *in vivo*, but which fail to cause a significant change from wild type in MST when inoculated individually in fish, are further examined. For these experiments, an inoculum dose of 10^7 CFU organisms are used, and animals are sacrificed at 4 and 8 weeks postinoculation. The liver, spleen, kidney, and/or other organs which are evident to one of skill in the art are harvested; one portion is homogenized for analysis of colony counts and another portion for histopathology.

Example 5 - Sequencing and characterizing regions flanking the transposons in the virulence mutants

Individual mutants confirmed in the goldfish model to be virulence mutants are examined by sequencing the nucleic acid flanking the site of insertion of the transposon. The sequence analysis can, of course, be performed before, simultaneously with, or after, a virulence defect has been confirmed.

A. Direct sequencing of flanking regions

In a most preferred embodiment, chromosomal DNA is isolated from each mutant and cut with a restriction enzyme that cuts once within the transposon (in this example, with *BamH1*). Linkers bearing a predefined PCR primer site, designed and generated using routine, art-recognized methods, are ligated to the BamH1-cut ends; and PCR fragments are amplified, using as primers a first outward primer sequence specific for a portion of the transposon, and a second inward primer specific for the PCR primer site in the appended linker, to generate an "amplified PCR fragment". In this example, a transposon-specific primer sequence is chosen based on the sequence of the inserted transposon, *IS 1096*. By "specific for," as used herein, is meant that a primer (*e.g.*, the first outward primer) is sufficiently complementary to a target (*e.g.*, the transposon) to bind to it (hybridize; serve as a PCR primer) under selected high stringent conditions, but not to bind to other, unintended, nucleic acids. Southern analysis, in which the membrane to which the DNA has been transferred is probed with an α-^{32}P labeled *aph* (kanamycin resistance) gene, can be used to identify the size of the "amplified PCR fragment" from each mutant. For example, mutants 41.2, 80.1 and 86.1 shown in Example 9 have unique amplified PCR fragments, of 550, 200 and 600 bp, respectively. The amplified PCR fragments are sequenced directly, using as primers one or both of the primers used to generate them, or are cloned into a vector such as pGEM and sequenced using primers corresponding to vector sequences. Methods for probing gels and sequencing DNA are routine and conventional in the art.

In another embodiment, the chromosomal DNA is cut with an enzyme which does not cut within the transposon. A variety of enzymes can be tested until one which generates a DNA fragment of an appropriate size is identified. Here, *Kpn* I is used. The

DNA is then ligated to create circular species and amplified by PCR using outward-facing primers complementary to the two ends of the transposon. In this way, the sequences which flank the insertion are amplified. These fragments are directly sequenced, using the same primers used to amplify the sequence.

B. Cloning and then sequencing flanking regions

In another embodiment, the gene sequences interrupted by a transposon are cloned first and then sequenced. Procedures for the analysis of DNA, including isolating DNA, cloning it, manipulating it, and sequencing it, are routine and well-known in the art. In a preferred embodiment, genomic DNA is extracted from each virulence mutant, and is digested with one or more restriction enzymes (*e.g.*, in this example, *Kpn*I or *Bam*HI) that provide genomic fragments of an appropriate size for cloning. The digested DNA is cloned into an appropriate plasmid, *e.g.*, Bluescript II KS (Promega), or a low-copy plasmid such as pACYC184, in *E. coli* DH5α, by using an appropriate positive selection marker *(e.g.*, kanamycin resistance). *Kpn*I does not cut within the transposon, so digestion with *Kpn* I, followed by selection with kanamycin, results in cloning of the transposon along with flanking DNA. *Bam* HI cuts once within the transposon, so digestion with *Bam* HI, followed by selection with kanamycin, results in cloning of part of the transposon along with flanking DNA on one side of the transposon. Once cloned, the gene sequence interrupted (disrupted) by the transposon is determined by using outward primers based on the sequence of the transposon insertion sequence, in this example, IS*1096* (See, *e.g.*, McAdam *et al* (1995). *Infec. Immun.* 63, 1004-1012).

C. Comparison of flanking sequences to known databases

DNA sequences flanking each transposon (localized on one or on both sides of the site of transposon insertion) are compared with the use of the BLAST programs provided in the National Center for Biotechnology Information (NCBI) data base.

In order to identify *M. tuberculosis* homologues of *M. marinum* virulence genes, the flanking sequences are also compared to the Mycobacterium database, using the advanced Blast search program, as above.

A discussion of functional homologues and related virulence genes from *M. tuberculosis* which have been identified for 3 *M. marinum* mutants is presented in Example 9.

Example 6 - Isolating and characterizing wild type *M. marinum* genes which correspond to the genes disrupted by transposons in avirulent *M. Marinum* mutants

Probes based on flanking *M. marinum* DNA sequences, characterized, *e.g.*, as in Example 5, are generated and used to screen an *M. marinum* cosmid library (The construction of such a cosmid library is described below). For example, part or all of the "amplified PCR fragment" which is described in Example 5 is labeled and used as a hybridization probe. Conditions for specifically hybridizing a probe to a target nucleic acid (*e.g.*, cosmid DNA) can be determined routinely by known methods in the art (see, *e.g.*, *Nucleic Acid Hybridization, a Practical Approach*, B.D. Hames and S.J. Higgins, eds., IRL Press, Washington, 1985). It is preferred that hybridization probing is done under selected high stringent conditions to ensure that the gene, and not a relative, is obtained. Of course, conditions of any stringency can be employed. By "high stringent" is meant that the gene hybridizes to the probe (*e.g.*, when the gene is immobilized on a filter) and the probe (which in this case is preferably about >200 nucleotides in length) is, *e.g.*, in solution, and the immobilized gene/hybridized probe is washed in 0.1X SSC at 65^0 C. for 10 minutes. SSC is 0.15M NaCl/0.015M Na citrate. In general, "high stringent hybridization conditions" are used which allow hybridization only if there are about 10% or fewer base pair mismatches. As used herein, "high stringent hybridization conditions" means any conditions in which hybridization will occur when there is at least 95%, preferably about 97 to 100%, nucleotide complementarity (identity) between the nucleic acids. The corresponding cosmid is identified; and individual virulence genes are subcloned from the cosmid clone, using routine, conventional procedures in the art. The complete gene sequence is determined by routine, conventional methods.

Construction of an *M. marinum* cosmid library: An *M. marinum* genomic library in an *E. coli* - Mycobacteria shuttle cosmid (pYUB18) is constructed, using, *e.g.*, methods disclosed in Jacobs, W.R. *et al* (1991). "Genetic Systems for Mycobacteria," in *Methods.*

Enzymol. 204, 537-555. The pYUB18 vector has a unique *Bam*HI site that can serve as the site of insertion of partial *Sau*3A-digested chromosomal DNA. Following *in vitro* packaging, the constructed libraries are transduced into cosmid *in vivo* packaging strains to permit amplification and efficient repackaging of recombinant cosmids into bacteriophage λ heads thus allowing for storage of the libraries as phage lysates.

Example 7 - Isolating and characterizing *M. tuberculosis* genes which correspond to *M. marinum* virulence genes

In order to identify an *M. tuberculosis* gene which corresponds to a particular *M. marinum* gene, an "amplified PCR fragment" from the *M. marinum* gene, such as that described in Example 5 or a fragment thereof, can be used to probe a cosmid library of *M. tuberculosis*. Most preferably, a probe based on the corresponding *M. tuberculosis* sequence, itself, is used. An *M. tuberculosis* cosmid library is constructed by routine methods. Hybridization is performed as described, *e.g.*, in Example 6. Positive cosmid clones are identified and the hybridizing sequences subcloned and sequenced, using routine, conventional, methods in the art.

Well-defined mutations can be introduced into a cloned *M. tuberculosis* gene, using the methods described herein for generating site-specific mutations in *M. marinum* genes. The mutations can then be introduced into the *M. tuberculosis* genome by homologous recombination. In a most preferred embodiment (as disclosed, *e.g.*, in Balasubramanian, V. *et al* (1996). *J. Bacteriol.* 178, 273-279, and Reyrat, J. *et al* (1995). *PNAS* 92, 8768-8772), the recombination is performed with long linear recombination substrates containing the mutated gene (virulence gene::*aph*) on a DNA fragment (>40 kb). This fragment is electroporated into the H37Rv strain of *M. tuberculosis* selecting for kanamycin resistance. Chromosomal DNA from the parent H37Rv strain and the kanamycin-resistant transformants are digested with *Kpn*I and probed with a *Kpn*I fragment containing the virulence gene::*aph* fragment. The strains containing the disrupted allele show a signal from a fragment which is 1.3-kb greater (*aph* gene) than the hybridizing fragment from the wild type gene clone (control). These mutant strains can

be tested, *e.g.*, in the guinea pig infection model (See, *e.g.*, Collins, D.M. *et al* (1995). *PNAS* 92, 8036-8040).

Alternatively, allelic exchange can be performed using *ts-sac*B vectors (see, *e.g.*, Pelicic *et al.* (1997). *PNAS* 94, 10955-10960). The virulence gene::*aph* construct is inserted into pJM10, a *ts-sac*B *E. coli* - *Mycobacteria* vector containing the kanamycin resistance gene for selection. The plasmid is introduced into the H37Rv strain of *M. tuberculosis* by electroporation with selection initially at 32°C on 7H10-kanamycin. Transformants are selected, grown in liquid culture, and then plated at 39°C on 7H10-kanamycin + 2% sucrose plates. Transformants obtained on the counterselective plates represent allelic exchange mutants.

Example 8 - Complementation assays

A candidate virulence gene is reintroduced into a transposon mutant on a low copy number *E. coli* - mycobacteria shuttle vector (pYUB213Δkm) (Ramakrishnan, L. *et al* (1997). *J. Bacteriol.* 179, 5862-5868) to determine whether the cloned gene complements the virulence defect in the goldfish model. This plasmid is a derivative of pMV262 (Stover, C.K. *et al* (1991). *Nature* 351, 456-460) with a bleomycin resistance gene for selection. Bacteria are recovered from those fish in which the virulence defect has been complemented, and analyzed for bleomycin and kanamycin resistance to confirm that the complementing plasmid is present.

Some cloned virulence gene candidates may fail to complement the virulence defect in the fish model because of, *e.g.*, instability of the cosmid clone, polar effects in the original mutation, requirement for a cluster of genes surrounding the interrupted gene, or toxic effects associated with overexpression of genes from multicopy plasmids. In order to overcome these problems, several alternative approaches can be used.

One approach is to utilize an integrating *E. coli* - mycobacterial shuttle vector, pMV361 (Stover, C.K. *et al* (1991). *Nature* 351, 456-460). The vector integrates in a site-specific manner into the chromosomal *attB* site. This site is in a well-conserved part of the mycobacterial genome and has been identified in BCG, *M. smegmatis, M. bovis, M.*

chelonei, M. leprae, M. phlei, and *M. tuberculosis*. Prior to the use of this vector in *M. marinum*, the presence of the *attB* site in *M. marinum* is confirmed by Southern blot analysis of *M. marinum* chromosomal DNA digested with *Bam*HI using a radiolabeled 1.7-kb *Sal* I *attB* fragment from *M. smegmatis*. In order to use this vector in mutants which contain the kanamycin resistance gene, the vector is modified to delete the kanamycin gene and to insert the bleomycin gene as was done, *e.g.*, with the construction of pYUB213Δkm (Ramakrishnan, L.H. *et al* (1997). *J. Bacter.* 179, 5862-5868). Using an integrating vector eliminates the possible instability seen with extrachromosomal plasmid maintenance *in vivo* (the integrated vector is stably maintained even without antibiotic selection), and the toxic effects associated with multicopy plasmids are reduced or eliminated since integration results in a single copy of the gene in the chromosome. To address the issue that the original transposon insertion phenotype was due to a polar effect on a downstream gene or that a cluster of genes is required for complementation, larger fragments of the original cosmid clone can be inserted into the integrating plasmid.

Another approach is to construct by allelic exchange specific chromosomal mutations in the identified virulence genes. Methods for using long linear recombination substrates for allelic exchange are provided, *e.g.*, in Balasubramanian, V. *et al* (1996). *J. Bacteriol.* 178, 273-279. Other methods for homologous recombination are found, *e.g.*, in Aldovini, A.R. *et al* (1993). *J. Bacteriol.* 175, 7282-7289; Norman, E. *et al* (1995). *Mol. Microbiol.* 16, 755-760; Baulard, A. *et al* (1996). *J. Bacteriol.* 178, 3091-3098; Marklund, B.I. *et al* (1995). *J. Bacteriol.* 177, 6100-6105; and Ramakrishnan, L. et al (1997). *J. Bacteriol.* 179, 5862-5868. These specific mutations allow the creation of non-polar mutations in the virulence genes.

Example 9 - Identification and characterization of thirteen *M. tuberculosis* virulence genes.

DNA regions flanking transposon insertion points for 13 mutants were amplified by inverse PCR and sequenced. Predicted amino acid sequences from all six reading frames of the DNA sequences obtained were subjected to similarity search of the nr database, using the NCBI BLAST program. The nr database includes, *e.g.*, all non-redundant GenBank CDS translations, PDB, SwissProt, PIR and PRF sequences. An

advanced BLAST search determined whether a homologous protein sequence was present in the *Mycobacterium tuberculosis* genome. The translated flanking sequences of mutants 41.2, 80.1, 86.1, 62.2, 67.1, 80.8, 39.2, 114.7, 32.2, 42.2, 60.2, 68.6 and 95.3 exhibited sequence identities with functionally homologous proteins from *M. tuberculosis* of 93%, 42%, 37-51%, 77%, 38%, 78%, 43%, 82%, 64%, 62%, 58-77%, 38%, and 36-47%, respectively.

Gene 41.2

The sequence of the flanking region of *M. marinum* mutant 41.2 is as follows:

5'-
CGGGCCGATCTATGACGAGNACGACGGGACAGATGGGTCCCCGGATGGTC
TA

CACCGAGACCAAACTGAACTCGTCGTTCTCCTTCGGCGGGCCCAAGTGTCT
GGTGAAGGTGATCCAAAAACTGTCCGGGTTGAGCATCAACCGGTTCATCGC
CATCGACTTCGTCGG - 3' (SEQ ID NO: 4)

This can be translated in the third reading frame to the following protein sequence:

1 GRSMTXTTGQ MGPRMVYTET KLNSSFSFGG PKCLVKVIQK LSGLSINRFI

51 AIDFV (SEQ ID NO: 5)

The mutant (41.2), when tested individually in the goldfish model, exhibits attenuated virulence as compared to the wild type organism (See Figure 8).

The gene interrupted in the attenuated mutant has been characterized by sequence analysis. Using the mycobacterium database, a functional homologue of this gene has been identified in *M. tuberculosis* (emb|CAA17628| (AL022004); (Rv0822c). Using the general genomic database, the gene has been shown to be most closely related to gene

emb|CAA20411|; (AL031317), a transcriptional regulator of *Streptomyces coelicolor* which belongs to the AraC family of transcriptional regulators. This suggests that the gene identified as interrupted in mutant 41.2 is a putative transcriptional regulator belonging to the AraC family.

The proteins belonging to this family have at least three main regulatory functions in common: carbon metabolism, stress response, and pathogenesis. (See, *e.g.*, Gallegos, M-T *et al* (1997). *Microbiology and Molecular Biology Reviews* 61, 393-410). Certain of these regulatory proteins are involved in the production of virulence factors in infections of plants or mammals. These regulatory factors have been found in microbes that colonize either the gastrointestinal, respiratory, or genitourinary tracts. These proteins are involved in stimulation of the synthesis of proteins that play a role in adhesion to epithelial tissues, components of the cell capsule, and invasins. Some members of the family control the production of other virulence factors. Some regulators are involved in the response to stressors, including oxidative stress and transition from exponential growth to the stationary phase. Without wishing to be bound by any mechanism, these observations suggest that the role of this gene in *M. tuberculosis* pathogenesis may be in invasion of the macrophage, survival in the macrophage (oxidative stress) or in transition to the latent state of tuberculosis (transition from exponential to stationary phase).

Gene 80.1

The sequence of the flanking region of *M. marinum* mutant 80.1 is as follows:

5' -

ACCTCCTGAATGTGTGACATGGCCCTAGAACCCTGCNTTAGACTATTTACAT
A

CATGGCTTCACCCGGCCGCCTGTGCCACTCATAAGACTACTGGAATGGACC
AACAATCGCACAGTCATCTGAAGCAGGAGTCTGTTAATCACAGGCCCTGAA
GGAACAGTGACTGTGCAGAGAAAGACGGCAATGCATCCTGTTAACTAAGT
GGCTGGAGGAGTGCCAGGTCATTCCAAAGAACATCCCTGAAATCTGGAGG
AGAAGGTATAGTGAGCACCCCAAAATTTCAACTGGAGACATCANACCAGA

GTCTCTACTGAGCTGCCAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACC
CCTTGGGGCCTCTAAACGGGTCTTGA - 3' (SEQ ID NO: 6)

This can be translated in the second reading frame to the following protein sequence:

1 PPECVTWP*N PALDYLHTWL HPAACATHKT TGMDQQSHSH LKQESVNHRP

51 *RNSDCAEKD GNASC*LSGW RSARSFQRTS LKSGGEGIVS TPKFQLETSX

QSLY*AAKLA AALE*LVNPL GPLNGS* (SEQ ID NO: 7)

The mutant (80.1), when tested individually in the goldfish model, exhibits attenuated virulence as compared to the wild type organism (See Figures 9 and 12).

The gene interrupted in the attenuated mutant has been characterized by sequence analysis, as described above for mutant 41.2. Functional homologues of this gene have been identified in *M. leprae* (sp|P54580|YV23 MYCLE; B2168 C2 209) and *M. tuberculosis* (sp|Q11162|YV23 MYCTU; CY20G9.23). Based on the sequence analysis, the gene identified as interrupted in mutant 80.1 is a hypothetical integral membrane protein, most closely related to a glutamate receptor channel, dbj|BAA02254.1 (D12822), from *Mus musculus*.

Gene 86.1

The sequence of the flanking region of *M. marinum* mutant 86.1 is as follows:

5'-TCATCGCTAACCGGTTGAGCTACCGCCCGCACAGCGTGCCCATCATCTC

CAACCTGACCGGCTCACTTGCCACAGTCGAGCAACTCACATCGCCCCGCTA
TTGGGCACAGCATGTACGGGAGCCAGTGCGGTTTCATGACGGCGTTACCGG
CTTGTTGGCAGGCGGAGAACA-3' (SEQ ID NO: 8)

This can be translated in the third reading frame to the following protein sequence:

1I A N R L S Y R P H S V P I I S N L T G S L A T V E Q L T S P R Y W A Q H V R E P V R F H D G V T G L L A G G E (SEQ ID NO: 9)

The mutant (86.1), when tested individually in the goldfish model, exhibits attenuation in virulence as compared to the wild type organism (See Figures 10 and 12).

The gene interrupted in the attenuated mutant has been characterized by sequence analysis, as described above for mutant 41.2. A family of functional homologues of this gene has been identified in *M. tuberculosis* (emb|CAB06094|Z83857|ppsE; emb|CAB06605|Z84725|pks6; emb|CAB09100|Z95617|pks9; emb|CAB09098|Z95617|pks8; emb|CAB06103|Z83858|pks1; pir|S73075|pks002c protein). Based on the sequence analysis, the gene identified as interrupted from mutant 86.1 is a polyketide synthase gene, most closely related to polyketide synthase genes AF263912 (*Streptomyces noursei*) and AF015823 (*Streptomyces venezuelae*).

Polyketides are lipid-like molecules that have potent biological activities. Examples of polyketides include antibiotics (erythromycin), immunosuppressants (rapamycin, FK506), antifungal agents (amphotericin B), antihelminthic agents (avermectin), and cytostatins (bafilomycin). A polyketide toxin has been recently described in *Mycobacterium ulcerans* (George, K.M. *et al* (1999). *Science* 283, 854-856) but no homologue was identified by sequence analysis in *M. tuberculosis*. Although it was recognized during analysis of the *M. tuberculosis* genome project that the genome contains a large number of polyketide synthesis genes, no polyketides from *M. tuberculosis* have been identified. That we have identified that a mutation in this gene attenuates the *M. marinum* strain in virulence suggests that although a polyketide toxin has not been identified, a product of this synthesis pathway is responsible for virulence. Without wishing to be bound to any mechanism, these observations suggest that a product of the polyketide synthesis pathway may be responsible for the tissue destruction and immunological modulation characteristic of diseases such as leprosy and tuberculosis.

Gene 62.2

The sequence of the flanking region of *M. marinum* mutant 62.2 is as follows:

GATCCGGTGCCGCCTTGACCGGCCGCGCCACCAGTACCGCCGACGCCGCCC
T G
GCCGCCGGCTTGTGCGGCTTGCGATGGGTCGGTGCTGTCGGTGCCGGTGCC
TCCGGTGCCGCCTTGGCCTCCGGTTCCGCCGGTGCCGCCCTGGCCGCCGGC
GCCTTGGATGCCGCCGGTGCCGGTTCCGGCTGCACCGCCCGTTCCGCCGGT
TCCGCCTGCGCCGCCGGTGCCT (SEQ ID NO: 10)

This can be translated in the -2 reading frame to the following protein sequence:

227 ggcaccggcggcgcaggcggaaccggcggaacgggcggtgcagcc

G T G G A G G T G G T G G A A

182 ggaaccggcaccggcggcatccaaggcgccggcggccagggcggc

G T G T G G I Q G A G G Q G G

137 accggcggaaccggaggccaaggcggcaccggaggcaccggcacc

T G G T G G Q G G T G G T G T

92 gacagcaccgacccatcgcaagccgcacaagccggcggccagggc

D S T D P S Q A A Q A G G Q G

47 ggcgtcggcggtactggtggcgcggccggtcaaggcggcaccgga

G V G G T G G A A G Q G G T G (SEQ ID NO: 12)

2 tc 1 (SEQ ID NO: 11)

The mutant (62.2), when tested individually in the goldfish model, exhibits attenuated virulence (reduced Competitive Index) as compared to the wild type organism (See Figure 12).

The gene interrupted in the attenuated mutant has been characterized by sequence analysis, as described above for mutant 41.2. Using either the mycobacterium or the general genomic database, a functional homologue of this gene has been identified in *M. tuberculosis* (emb|CAA17748.1| (AL022022); (Rv3511).

This is a hypothetical glycine-rich protein (Rv3511) belonging to a large *M. tuberculosis* PE- PGRS protein family, which comprises roughly 5% of the coding DNA of *M. tuberculosis*. The genes of this family are scattered throughout the genome of *M. tuberculosis* and other closely related mycobacteria. This family is characterized by a relatively conserved amino acid NH_2-terminus. The function of these proteins is unknown but some hypotheses are that they represent a source of antigenic diversity or that their glycine repeats inhibit host major histocompatibility complex class I processing, akin to the glycine repeats of the Epstein-Barr virus EBNA-1 protein. That we have identified that a mutation in this gene attenuates the *M. marinum* strain in virulence suggests that the protein product of this gene is responsible for the immunological modulation characteristic of diseases such as leprosy and tuberculosis.

Gene 67.1

The sequence of the flanking region of *M. marinum* mutant 67.1 is as follows:

GGTCGAAGACTATCGGTATGCTCCATAGCGTTCCGTCGGGAAGCTGCATGT
TGTCAAGGGTTTCGTCGACCTCTCGGCGACCCATGAATCCCGATAGTGGCG
TGAAGAAACCGTACGAGATGCTGATCACCTCGTGGGCGGTCGCCTTCGATA
TCGGGATGCGCACCAATCCCTCAATCCGGCCGGCCACGTTTTCCCTTTCCAC
CCTGTCGACGAGTGGGTGTCCGTTATGGCCTAAATAATCCATCTTGCTGCCT
CTTTCTGAAATCGAATTTATTACTATCG (SEQ ID NO: 13)

This can be translated in the six reading frames to the following protein sequences:

DNA: GGTCGAAGACTATCGGTATGCTCCATAGCGTTCCGTCGGGAAGCTGCATGT

+3: S K T I G M L H S V P S G S C M L

+2: V E D Y R Y A P * R S V G K L H V

+1: G R R L S V C S I A F R R E A A C

DNA: TGTCAAGGGTTTCGTCGACCTCTCGGCGACCCATGAATCCCGATAGTGGCG

+3: S R V S S T S R R P M N P D S G V

+2: V K G F V D L S A T H E S R * W R

+1: C Q G F R R P L G D P * I P I V A

DNA: TGAAGAAACCGTACGAGATGCTGATCACCTCGTGGGCGGTCGCCTTCGATA

+3: K K P Y E M L I T S W A V A F D I

+2: E E T V R D A D H L V G G R L R Y

+1: * R N R T R C * S P R G R S P S I

DNA: TCGGGATGCGCACCAATCCCTCAATCCGGCCGGCCACGTTTTCCCTTTCCA

+3: G M R T N P S I R P A T F S L S T

+2: R D A H Q S L N P A G H V F P F H

+1: S G C A P I P Q S G R P R F P F P

DNA: CCCTGTCGACGAGTGGGTGTCCGTTATGGCCTAAATAATCCATCTTGCTGC

+3: L S T S G C P L W P K * S I L L P

+2: P V D E W V S V M A * I I H L A A

+1: P C R R V G V R Y G L N N P S C C

DNA: CTCTTTCTGAAATCGAATTTATTACTATCG (SEQ ID NO: 13)

+3: L S E I E F I T I (SEQ ID NO: 14)

+2: S F * N R I Y Y Y (SEQ ID NO: 15)

+1: L F L K S N L L L S (SEQ ID NO: 16)

DNA: CGATAGTAATAAATTCGATTTCAGAAAGAGGCAGCAAGATGGATTATTTAG

-1: R * * * I R F Q K E A A R W I I *

-2: D S N K F D F R K R Q Q D G L F R

-3: I V I N S I S E R G S K M D Y L G

DNA: GCCATAACGGACACCCACTCGTCGACAGGGTGGAAAGGGAAAACGTGGCCG

-1: A I T D T H S S T G W K G K T W P

-2: P * R T P T R R Q G G K G K R G R

-3: H N G H P L V D R V E R E N V A G

DNA: GCCGGATTGAGGGATTGGTGCGCATCCCGATATCGAAGGCGACCGCCCACG

-1: A G L R D W C A S R Y R R R P P T

-2: P D * G I G A H P D I E G D R P R

-3: R I E G L V R I P I S K A T A H E

DNA: AGGTGATCAGCATCTCGTACGGTTTCTTCACGCCACTATCGGGATTCATGG

-1: R * S A S R T V S S R H Y R D S W

-2: G D Q H L V R F L H A T I G I H G

-3: V I S I S Y G F F T P L S G F M G

DNA: GTCGCCGAGAGGTCGACGAAACCCTTGACAACATGCAGCTTCCCGACGGAA

-1: V A E R S T K P L T T C S F P T E

-2: S P R G R R N P * Q H A A S R R N

-3: R R E V D E T L D N M Q L P D G T

DNA: CGCTATGGAGCATACCGATAGTCTTCGACC (SEQ ID NO: 17)

-1: R Y G A Y R * S S T (SEQ ID NO: 18)

-2: A M E H T D S L R (SEQ ID NO: 19)

-3: L W S I P I V F D (SEQ ID NO: 20)

The mutant (67.1), when tested individually in the goldfish model, exhibits attenuated virulence as compared to the wild type organism (See Figures 12 and 13).

The gene interrupted in the attenuated mutant has been characterized by sequence analysis, as described above for mutant 41.2. Using the mycobacterium database, a functional homologue of this gene has been identified in *M. tuberculosis* (emb|CAB08565.1|(Z95324) purA. This homologue, in the +2 frame, with an identity 38% (similarity of 57%), is an adenylosuccinate synthetase *(M. tuberculosis* homologue 008381). This protein product plays an important role in the de novo pathway of purine nucleotide biosynthesis. Thus in the host animal, particularly in the macrophage where

nutrients may be limiting the product of this gene may be required for survival of *Mycobacterium marinum* and *M. tuberculosis*.

Based on the sequence analysis to the entire genomic database, the gene identified as interrupted from mutant 67.1 is a sulfate adenylyltransferase with homology to diverse organisms including *Pyrococcus abyssi, Synechocystis* sp., and *Bacillus subtilis*. The homology is in the –3 reading frame of the translated gene product and shows 27-40% identity (51-62% similar). The homology noted to the sulfate adenylyltransferase enzymes suggests that mutant 67.1 is attenuated in its ability to respond to sulfate starvation as this enzyme is required for growth in defined synthetic medium with sulfate as a sulfur source. This suggests that in the animal host a sulfur source is limiting and thus interruption of this gene attenuates growth of the organism in the animal host. Thus interruption of this gene in a live attenuated Mycobacterium vaccine strain would be beneficial, as it will limit the ability of the vaccine strain to grow in the animal host.

Gene 80.8

The sequence of the flanking region of *M. marinum* mutant 80.8 is as follows:

CCAATTAGCTGATTATTCCTCGGGCGTGCTCAACGCCAAGGACTACATATC
AGGTTACTTCCACTAAAATTCGCGGGCCCCGATCGGCGACATTACTCGACG
GTTTTCGGGGGAATCTCAGCGGTGATGGCATTCTTGAGGGCGACGTAGCGT
TTGGCGTCGGGATC (SEQ ID NO: 21)

This can be translated in the -1 reading frame to the following protein sequence:

DPDAKRYVALKNAITAEIPPKTVE*CRRSGPANFSGSNLICSPWR*ARPR

NNQLI (SEQ ID NO: 22)

The mutant (80.8), when tested individually in the goldfish model, exhibits attenuation in virulence (reduced Competitive Index) as compared to the wild type organism (See Figure 12).

The gene interrupted in the attenuated mutant has been characterized by sequence analysis, as described above for mutant 41.2. Using either the mycobacterium or or the general genomic database, a functional homologue of this gene has been identified in *M. tuberculosis* (emb|CAB02482.1|Z80343|lipE. This is a probable carboxylic-ester hydrolase (*M. tuberculosis* homologue Rv3775) also referred to as an esterase or *lipE*. The homology is in the –1 reading frame with 83% similarity, 78% identity. This gene may have a role in fatty acid synthesis in Mycobacterium species or may be involved in establishment or dissemination in the animal host by destruction of the host cell fatty acids present in the host cell membrane. That we have identified that a mutation in this gene attenuates the *M. marinum* strain in virulence suggests that the protein product of this gene is responsible for the virulence attributes of Mycobacterium species and may contribute to the establishment of diseases such as leprosy and tuberculosis.

Gene 39.2

The sequence of the flanking region of *M. marinum* mutant 39. is as follows:

GATCCGCTGGACGGCACCAAAGAATTCATCAAGGGCAGCGATGAGTTCAC
CGTCAACATCGCCCTGGTCGAGAACCAGGAACCCATTCTCGGGGCAATCTA
CGGTCCAGCGAAGCAACTTCTGCACTACGCGGCCAAAGGGGCT (SEQ ID NO: 23)

This can be translated in the +1 reading frames to the following protein sequence:

7 ctggacggcaccaaagaattcatcaagggcagcgatgagttcacc

L D G T K E F I K G S D E F T

52 gtcaacatcgccctggtcgagaaccaggaacccattctcggggca

V N I A L V E N Q E P I L G A

97 atctacggtccagcgaagcaacttctgcactacgcggccaaaggg

I Y G P A K Q L L H Y A A K G

142 gct 144 (SEQ ID NO: 43)

A (SEQ ID NO: 24)

The mutant (39.2), when tested individually in the goldfish model, exhibits attenuation in virulence as compared to the wild type organism (See Figure 14).

The gene interrupted in the attenuated mutant has been characterized by sequence analysis, as described above for mutant 41.2. Using the mycobacterium database, a functional homologue of this gene has been identified in *M. tuberculosis* (emb|CAB06277.1|Z8386|hypothetical protein Rv3137). This homologue, in the +1 frame, with an identity 43% (similarity of 63%), is a probable inositol monophosphate phosphatase, because it contains an inositol monophosphatase family signature sequence. It is related to the cysQ proteins identified in the whole database search described below, which also belong to the inositol monophosphatase family.

Based on a sequence analysis to the entire genomic database, the gene identified as interrupted from mutant 39.2 is predicted to be a structural protein of an ammonium transport system (also known as a cysQ gene). This protein affects the pool of 3'-phosphoadenosine -5'-phosphosulfate in the pathway of sulfite synthesis. The identity is in the +1 reading frame of the translated gene product and is 53-65% identical (63-82% similar). The homology noted suggests that mutant 39.2 is attenuated in its ability to respond to sulfate starvation as this enzyme is required for growth in defined synthetic medium with sulfate as a sulfur source. This suggests that in the animal host a sulfur source is limiting and thus interruption of this gene attenuates growth of the organism in the animal host. Thus interruption of this gene in a live attenuated Mycobacterium vaccine

strain would be beneficial, as it will limit the ability of the vaccine strain to grow in the animal host.

Gene 114.7

The sequence of the flanking region of *M. marinum* mutant 114.7 is as follows:

AGCCGTATTTCGCCATTGAGAGTTGGGGTCTTGAGATCGGCACTGGAAGGG
GACAGCGTGCTATTGCCTCTTGGTCCGCCCTTGCCACCTGATGCTGTGGCGG
CTAAACGGGGTGAGTCGGGGCTGCTCTGCGGCTTGTCGGTTCCGCTCAGCT
GGGGTACGGCCGTTCCGCCGGATGACTACNACCATTGGGCACCGGAGCCTG
AAGAAGGCGCCGAGGCCGTGGTCGAAGAAAACGTGGATGCGGCAGCTGCC
GGTACCGACGAGTGGGACGAGTGGGCGGAATGGAGGGAGTGGGAGGCAG
CAAATGCCCGAACCTCATTTTCGAGATGCCCCGTACCAGCAGCCGTGATAC
CCGAACTCGCCGGCGGCCGGTTGAGA (SEQ ID NO: 25)

This can be translated in the +1 reading frames to the following protein sequence:

16 ttgagagttggggtcttgagatcggcactggaaggggacagcgtg

L R V G V L R S A L E G D S V

61 ctattgcctcttggtccgcccttgccacctgatgctgtggcggct

L L P L G P P L P P D A V A A

106 aaacggggtgagtcggggctgctctgcggcttgtcggttccgctc

K R G E S G L L C G L S V P L

151 agctggggtacggccgttccgccggatgactacnaccattgggca

S W G T A V P P D D Y X H W A

196 ccggagcctgaagaaggcgccgaggccgtggtcgaagaaaacgtg

P E P E E G A E A V V E E N V

241 gatgcggcagctgccggtaccgacgagtgggacgagtgggcggaa

D A A A A G T D E W D E W A E

286 tggagggagtgggaggcagcaaatgcccgaacctcattttcgaga

W R E W E A A N A R T S F S R

331 tgccccgtaccagcagccgtgatacccgaactcgccggcggccgg

C P V P A A V I P E L A G G R

376 ttgaga 381 (SEQ ID NO: 44)

L R (SEQ ID NO: 26)

The mutant (114.7), when tested in pools in the goldfish model, appears to exhibit attenuation in virulence as compared to the wild type organism.

The gene interrupted in the attenuated mutant has been characterized by sequence analysis. Using either the mycobacterium or the general genomic database, a functional homologue of this gene has been identified in *M. tuberculosis* (pir E70662); (Rv2348c). The homology is in the +1 reading frame, with an identity of 82% (similarity 84%), to a hypothetical protein of *M. tuberculosis*. This protein is of unknown function as it has no known homology to any other sequence in the database. Extrapolating from the animal model, it appears that this gene is a virulence gene in *M. marinum* and *M. tuberculosis*.

Mutant 32.2

The sequence of the flanking region of *M. marinum* mutant 32.2 is as follows:

TCCANNCAGAGGNGCACGTAGANCGTAGGACGGAANGCGGNGNGATCGNC
AATACGGCTGGCNCTGCNAGAACTGNTCGAGGGCCTGCNGCTGGGGCC

(SEQ ID NO: 27)

This can be translated in the –2 reading frame to the following protein sequence:

APAAGPRXVLAXPAVLXIXPXSVLRSTCXSXW (SEQ ID NO: 28)

The mutant 32.2, when tested individually in the goldfish model, exhibits attenuated virulence (reduced Competitive Index, see Figure 12) as compared to the wild type organism.

The gene interrupted in the attenuated mutant has been characterized by sequence analysis. Using the Mycobacterium database, a functional homologue of this gene has been identified in *M. tuberculosis* (emb CAB06230 (Z83864) (Rv3860). This is a gene encoding a hypothetical protein of unknown function with homology to other Mycobacterium proteins also of unknown function including [emb CAB08086 (Z94121) (Rv3888c); emb CAA75199 (Y14967); emb CAA17968 (AL022120) (Rv3876); emb CAB08981 (Z95558) (Rv0530) and emb CAA15582 (AL008967) (Rv2787)]. That we have identified that a mutation in this gene attenuates the *M. marinum* strain in virulence suggests that the protein product of this gene contributes to the disease process in tuberculosis and leprosy. The interruption of this gene in a live attenuated Mycobacterium vaccine strain would be beneficial, as it will limit the ability of the vaccine strain to grow in the animal host.

The homology with the *M. tuberculosis* homologue is 64% identity, 78% similarity.

Mutant 42.2

The sequence of the flanking region of *M. marinum* mutant 42.2 is as follows:

TTTGCAATCCACCTGTACGCGGAACTNTTNANNNCCGTTTTGCCTTGNCGA
ATAAGCTAGCT (SEQ ID NO: 29)

This can be translated in the –1 reading frame to the following protein sequence:

S*LIRQGKTXXXSSAYRWIA (SEQ ID NO: 30)

The mutant 42.2, when tested individually in the goldfish model, exhibits attenuated virulence (reduced Competitive Index, see Figure 12 and decreased virulence in LD50 experiment, Figure 15) as compared to the wild type organism.

The gene interrupted in the attenuated mutant has been characterized by sequence analysis. Using the Mycobacterium database, a functional homologue of this gene has been identified in *M. tuberculosis* (emb CAB03756 (Z81371) (mbtB). This is a gene involved in mycobactin biosynthesis. *M. tuberculosis* produces both cell associated mycobactins and secreted, water-soluble mycobactins. Both types are siderophores and act to scavenge iron from the environment to support growth of the organism. The genes involved in mycobactin synthesis are contained in an operon. That we have identified that a mutation in this gene attenuates the *M. marinum* strain in virulence suggests that iron is required for *Mycobacterium* growth in the animal host. The interruption of this gene in a live attenuated Mycobacterium vaccine strain would be beneficial, as it will limit the ability of the vaccine strain to grow in the animal host.

The homology with the *M. tuberculosis* homologue is 62% identity, 99% similarity.

Mutant 60.2

The sequence of the flanking region of *M. marinum* mutant 60.2 is as follows:

CCANACCTATCTGTTTNCAGNTTNAGACNACGGNATCTCACGCGNTTGGGC
CCNGCCACCAAACGCCGCGTNGA (SEQ ID NO: 31)

This can be translated in six reading frames to the following protein sequences:

DNA: CCANACCTATCTGTTTNCAGNTTNAGACNACGGNATCTCACGCGNTTGGGC

+3: X P I C X Q X X T T X S H A X G P

+2: X T Y L F X X X D X G I S R X W A

+1: P X L S V X X X R X R X L T R L G

DNA: CCNGCCACCAAACGCCGCGTNGA (SEQ ID NO: 31)

+3: X H Q T P R X (SEQ ID NO: 32)

+2: X P P N A A X (SEQ ID NO: 33)

+1: P A T K R R V (SEQ ID NO: 34)

>60.2/T89 T87 removed

DNA: TCNACGCGGCGTTTGGTGGCNGGGCCCAANCGCGTGAGATNCCGTNGTCTN

-1: S T R R L V A G P X R V R X R X L

-2: X R G V W W X G P X A * D X V V X

-3: X A A F G G X A Q X R E X P X S X

DNA: AANCTGNAAACAGATAGGTNTGG (SEQ ID NO: 35)

-1: X L X T D R X (SEQ ID NO: 36)

-2: X X K Q I G X (SEQ ID NO: 37)

-3: X X N R * V W (SEQ ID NO: 38)

The mutant 60.2, when tested individually in the goldfish model, exhibits attenuated virulence (reduced Competitive Index, see Figure 12) as compared to the wild type organism.

The gene interrupted in the attenuated mutant has been characterized by sequence analysis. Using the Mycobacterium database, functional homologues of this gene have been identified in *M. tuberculosis* [emb CAA17485 (AL021957) (Rv2181); emb CAB06507 (Z84498) (Rv1954c); emb CAA17586 (AL021999) (Rv0987); emb CAB07087 (Z92771) (Rv3268); emb CAB08632 (Z95387) (Rv2610c)]. This is a gene encoding a hypothetical integral membrane protein of unknown function. That we have identified that a mutation in this gene attenuates the *M. marinum* strain in virulence suggests that it is required for *Mycobacterium* growth in the animal host. The interruption of this gene in a live attenuated Mycobacterium vaccine strain would be beneficial, as it will limit the ability of the vaccine strain to grow in the animal host.

The homology with the *M. tuberculosis* homologue Rv 2181 is 58% identity, 66% similarity, overall homology with all the genes identified is 58-77% identity, 66-88% similarity.

Gene 68.6

The sequence of the flanking region of *M. marinum* mutant 68.6 is as follows:

AAATCATCATCTATCGTTACCCGGGGCAAGCCAAGCACCTCAGCAAAAATT
CTGCAGAGCATTTCCTCTTGCGGAGTTCGCGGCATACGGCCAATCGCCGCA
TGATGATCGGGCACAGGCAGCGCTTTACGATCCACCTTCTTATTCGGAGTT
AACGGCATGGTCTCAAGTCTTACGATGACAGACGGCACCATATATTCGGCC
AGTTTCAGGGAGGCGTAGCGCCGCAGTTCTGCTGTATCTATCA

(SEQ ID NO: 39)

This can be translated in the -3 reading frame to the following protein sequence:

1 IDTAELRRYA SLKLAEYMVP SVIVRLETMP LTPNKKVDRK ALPVPDHHAA IGRMPRTPQE EMLCRIFAEV LGLPRVTIDD D (SEQ ID NO: 40)

The mutant (68.6), when tested individually in the goldfish model, exhibits attenuated virulence (reduced Competitive Index) as compared to the wild type organisms (Figure 12).

The gene interrupted in the attenuated mutant has been characterized by sequence analysis. Using the mycobacterium database, a functional homologue of this gene has been identified in *M. tuberculosis* (pir E70751 emb CAA98937 Z74410); (nrp protein). The homology is in the -3 reading frame, with an identity of 43% (similarity 62%), to a probable nrp protein of *M. tuberculosis*. This protein belongs to a superfamily of acetate CoA ligase proteins involved in peptide synthesis. A second protein of *M. tuberculosis* also shows significant homology. This protein is the mbtE protein (pir C70588 emb CAB08481 Z95208). The homology is again in the –3 reading frame, with an identity of 38% (similarity 56%). This is a gene involved in mycobactin biosynthesis. *M. tuberculosis* produces both cell associated mycobactins and secreted, water-soluble mycobactins. Both types are siderophores and act to scavenge iron from the environment to support growth of the organism. The genes involved in mycobactin synthesis are contained in an operon.

Searching against the entire database, we have identified significant homologues in *Bacillus subtilis*. The gene homologue is *dhbF* a gene encoding the 2,3-dihydroxybenzoate biosynthesis. The gene has been identified as essential for the synthesis of a siderophore in *B. subtilis*.

Mutant 95.3

The sequence of the flanking region of *M. marinum* mutant 95.3 is as follows:

GATTAGCTTATTCCTCAAGGCACGAGCGATTAGCTTATTCCTCAAGGCACG AGCGACTAGCTTATTCCTCAAGGCACGAGCTTCGCACTTGACGGTGTAGAG

CTCAATAGCTTATTCCTCAAGGCACGAGCTCGACTTCGCACTTGACGGTGT
AGAGCTCAAAG (SEQ ID NO: 41)

This can be translated in the +1 reading frame to the following protein sequence:

1 D*LIPQGTSD*LIPQGTSD*LIPQGTSFALDGVELNSLFLKARARLRT*R

52 CRAQ (SEQ ID NO: 42)

The gene interrupted in the attenuated mutant has been characterized by sequence analysis. Using the Mycobacterium database, functional homologues of this gene have been identified in *M. tuberculosis* [pir B70963 emb CAB0717 (Z92669) (Rv0236c); pir B70748 emb CAA98982 (Z74697) smc protein]. This is a gene encoding a hypothetical integral membrane protein of unknown function. That we have identified that a mutation in this gene attenuates the *M. marinum* strain in virulence suggests that it is required for *Mycobacterium* growth in the animal host. The interruption of this gene in a live attenuated Mycobacterium vaccine strain would be beneficial, as it will limit the ability of the vaccine strain to grow in the animal host.

The homology with the *M. tuberculosis* homologue Rv 0236c is 36% identity, 64% similarity and with the smc protein is 47% identity, 61% similarity.

From the foregoing description, one skilled in the art can easily ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make changes and modifications of the invention to adapt it to various usage and conditions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The preceding preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The entire disclosure of all applications, patents and publications, cited above and in the figures are hereby incorporated by reference.

We claim:

1. A method for identifying a virulence gene of *M. marinum*, comprising

a) mutagenizing an *M. marinum* bacterium by introducing into the bacterium a plasmid which comprises a signature-tagged transposon, whereby the transposon integrates into and disrupts a gene in the bacterium,

b) introducing the mutagenized bacterium into a host susceptible to infection thereof,

c) identifying a bacterium which comprises a signature tagged transposon and which exhibits reduced viability in the host, compared to a non-mutagenized *M. marinum* bacterium,

d) cloning and/or sequencing a nucleic acid sequence which flanks the integrated transposon in said identified bacterium, and

e) identifying a wild type *M. marinum* gene which comprises at least a portion of said flanking sequence.

2. The method of claim 1, further comprising

f) confirming that the mutation renders *M. marinum* less virulent.

3. A method of constructing an avirulent *M. marinum* bacterium, comprising mutagenizing an *M. marinum* virulence gene identified by the method of claim 1.

4. An avirulent *M. marinum* bacterium, produced by the method of claim 3.

5. An avirulent *M. marinum* bacterium, in which one or more genes comprising a nucleic acid of SEQ ID NOs: 4, 6, 8, 10, 11, 13, 21, 23, 25, 27, 29, 31, 35, 39, 41, 43 or 44 is mutated.

6. A method for identifying a virulence gene of *M. tuberculosis*, comprising identifying a virulence gene of *M. marinum* bacterium according to the method of claim 1, and further comprising,

comparing said flanking nucleic acid sequence to a databank of *M. tuberculosis* nucleic acid sequences, and/or comparing the sequences of peptides which are coded for by said flanking sequences to a known *M. tuberculosis* protein database, and

identifying an *M. tuberculosis* gene which comprises a sequence that is substantially identical to said flanking sequences.

7. A method for generating an avirulent *M. tuberculosis* bacterium, comprising mutagenizing an *M. tuberculosis* virulence gene identified by the method of claim 6.

8. An avirulent *M. tuberculosis* bacterium, produced by the method of claim 7.

9. An avirulent *M. tuberculosis* bacterium, in which one or more of genes Rv0822c, CY20G9.23 (Rv0497), the pks family, including e.g., ppsE (Rv2935), psk6 (Rv0405), pks9 (Rv1664), pks8 (Rv1662), pks1 (Rv2946c), and pks002c, Rv3511, O08381 (Rv0357c), Rv3775, Rv3137, Rv2348c, Rv3860, mbtB (Rv2383c), Rv2181, Rv1954c, Rv0987, Rv3268, Rv2610c, nrp (pir E70751, Rv0101), mbtE (Rv2380c), Rv0236c or smc (Rv2922c) is mutated to render the *M. tuberculosis* bacterium less virulent.

10. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene Rv0822c is mutated.

11. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene CY20G9.23 is mutated.

12. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene ppsE is mutated.

13. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene pks6 is mutated.

14. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene pks9 is mutated.

15. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene pks8 is mutated.

16. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene pks1 is mutated.

17. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene pks002c is mutated.

18. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene Rv3511 is mutated.

19. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene 008381 is mutated.

20. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene Rv3775 is mutated.

21. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene Rv3137 is mutated.

22. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene Rv2348c is mutated.

23. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene Rv3860 is mutated.

24. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene mbtB is mutated.

25. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene Rv2181, Rv1954c, Rv0987, Rv3268, or Rv2610c is mutated.

26. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene nrp (pirE 70751) or mbtE is mutated.

27. An avirulent *M. tuberculosis* bacterium of claim 9, in which gene Rv0236c or smc is mutated.

28. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 4, or a variant or fragment thereof.

29. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 6, or a variant or fragment thereof.

30. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 8, or a variant or fragment thereof.

31. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 11, or a variant or fragment thereof.

32. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 13, or a variant or fragment thereof.

33. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 21, or a variant or fragment thereof.

34. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 23, or a variant or fragment thereof.

35. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 25, or a variant or fragment thereof.

36. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO:27, or a variant or fragment thereof.

37. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 29, or a variant or fragment thereof.

38. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 31, or a variant or fragment thereof.

39. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 39, or a variant or fragment thereof.

40. An isolated nucleic acid comprising the oligonucleotide of SEQ ID NO: 41, or a variant or fragment thereof.

41. A pharmaceutical composition, comprising an avirulent *M. marinum* bacterium of claim 5 and a pharmaceutically acceptable carrier.

42. An attenuated *M. marinum* vaccine, comprising an avirulent *M. marinum* bacterium of claim 5 and a pharmaceutically acceptable carrier.

43. A pharmaceutical composition, comprising an avirulent *M. tuberculosis* bacterium of claim 9 and a pharmaceutically acceptable carrier.

44. An attenuated *M. tuberculosis* vaccine, comprising an avirulent *M. tuberculosis* bacterium of claim 9 and a pharmaceutically acceptable carrier.

45. An attenuated *M. tuberculosis* vaccine, comprising an avirulent *M. tuberculosis* bacterium which comprises one or more mutations in one or more virulence genes identified by the method of claim 7 and a pharmaceutically acceptable carrier.

46. A method to elicit an immune response in a fish in need of such treatment, comprising administering to said fish an avirulent *M. marinum* bacterium of claim 5.

47. A method to elicit an immune response in a patient in need of such treatment, comprising administering to said patient an avirulent *M. tuberculosis* bacterium of claim 9.

48. An isolated polyketide made by the polyketide synthase encoded by the *M. marinum* polyketide synthase gene which comprises the oligonucleotide of SEQ ID NO:8.

49. An isolated polyketide made by the *M. tuberculosis* polyketide synthase gene ppsE, pks6, pks8, pks9, pks1 or pks002c .

50. A method for isolating a mutagenized *M. marinum* bacterium which exhibits reduced virulence in a host susceptible to infection thereof compared to a non-mutagenized *M. marinum* bacterium, comprising integrating a tagged transposon into the DNA of a *M. marinum* bacterium in a manner effective to produced reduced virulence, and isolating said mutagenized bacterium.

51. An isolated nucleic acid of claim 28, consisting essentially of the oligonucleotide of SEQ ID NO: 4, or a variant or fragment thereof.

52. An isolated nucleic acid of claim 29, consisting essentially of the oligonucleotide of SEQ ID NO: 6, or a variant or fragment thereof.

53. An isolated nucleic acid of claim 30, consisting essentially of the oligonucleotide of SEQ ID NO: 8, or a variant or fragment thereof.

54. An isolated nucleic acid of claim 31, consisting essentially of the oligonucleotide of SEQ ID NO: 11, or a variant or fragment thereof.

55. An isolated nucleic acid of claim 32, consisting essentially of the oligonucleotide of SEQ ID NO: 13, or a variant or fragment thereof.

56. An isolated nucleic acid of claim 33, consisting essentially of the oligonucleotide of SEQ ID NO: 21, or a variant or fragment thereof.

57. An isolated nucleic acid of claim 34, consisting essentially of the oligonucleotide of SEQ ID NO: 23, or a variant or fragment thereof.

58. An isolated nucleic acid of claim 35, consisting essentially of the oligonucleotide of SEQ ID NO: 25, or a variant or fragment thereof.

59. An isolated nucleic acid of claim 36, consisting essentially of the oligonucleotide of SEQ ID NO: 27, or a variant or fragment thereof.

60. An isolated nucleic acid of claim 37, consisting essentially of the oligonucleotide of SEQ ID NO: 29, or a variant or fragment thereof.

61. An isolated nucleic acid of claim 38, consisting essentially of the oligonucleotide of SEQ ID NO: 31, or a variant or fragment thereof.

62. An isolated nucleic acid of claim 39, consisting essentially of the oligonucleotide of SEQ ID NO: 39, or a variant or fragment thereof.

63. An isolated nucleic acid of claim 40, consisting essentially of the oligonucleotide of SEQ ID NO: 41, or a variant or fragment thereof.

64. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 51.

65. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 52.

66. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 53.

67. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 54.

68. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 55.

69. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 56.

70. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 57.

71. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 58.

72. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 59.

73. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 60.

74. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 61.

75. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 62.

76. An isolated nucleic acid which is complementary to, or which can hybridize under high stringency conditions to, at least a portion of the isolated nucleic acid, or a variant thereof, of claim 63.

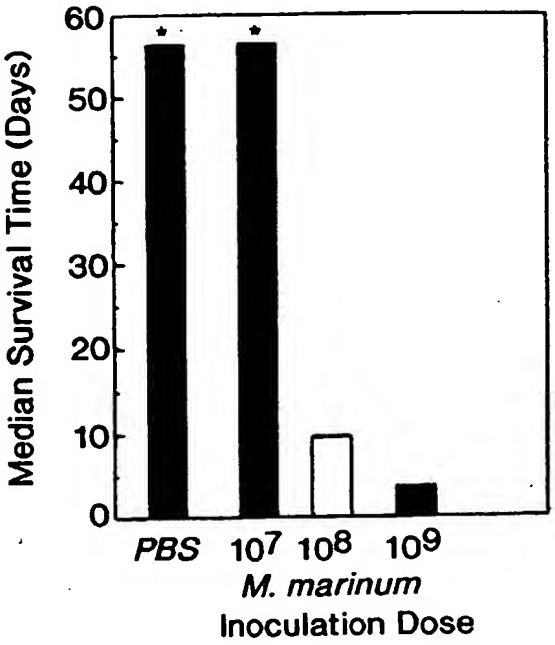
Median Survival Time (Days)
60
50
40
30
20
10
0
PBS
10^7
10^8
10^9
M. marinum
Inoculation Dose

FIG. 1

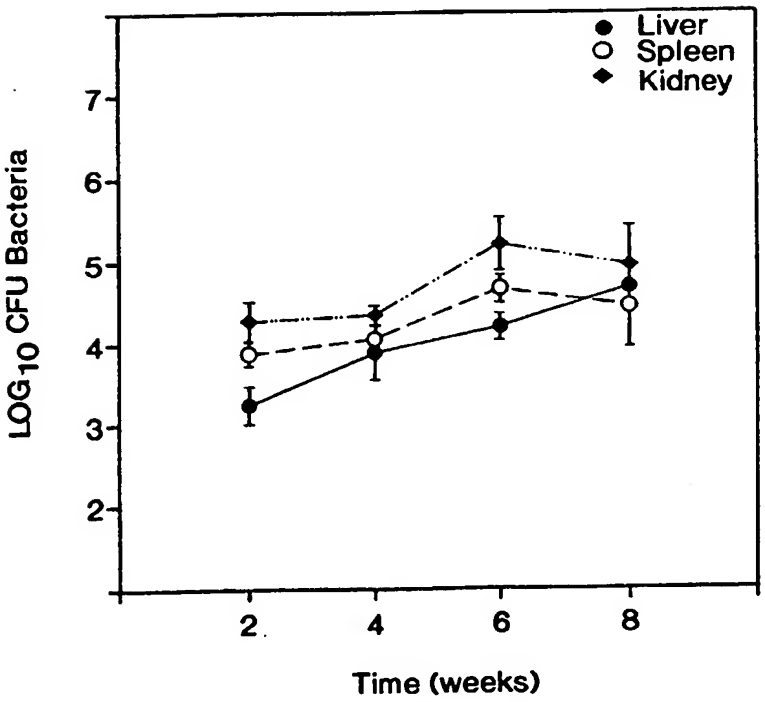
Liver
Spleen
Kidney
LOG10 CFU Bacteria
7
6
5
4
3
2
2
4
6
8
Time (weeks)

FIG. 2

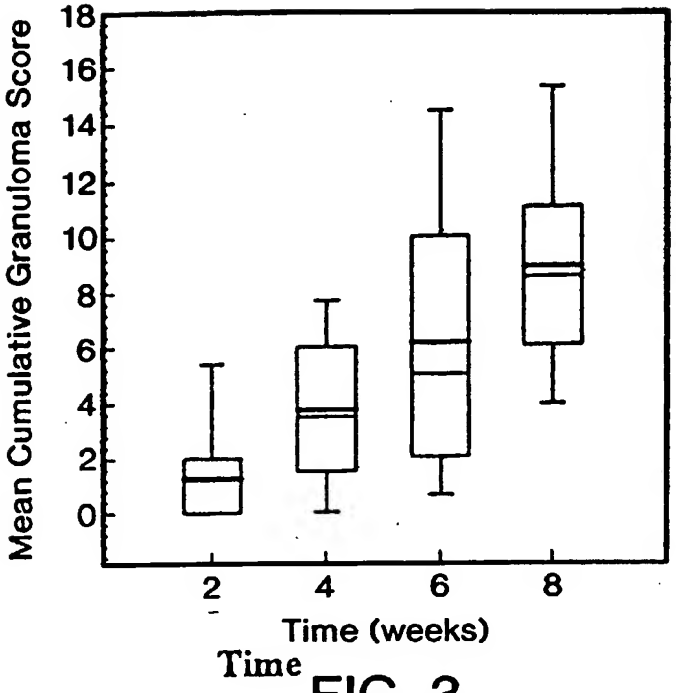
Mean Cumulative Granuloma Score
18
16
14
12
10
8
6
4
2
0
2
4
6
8
Time (weeks)
Time

FIG. 3

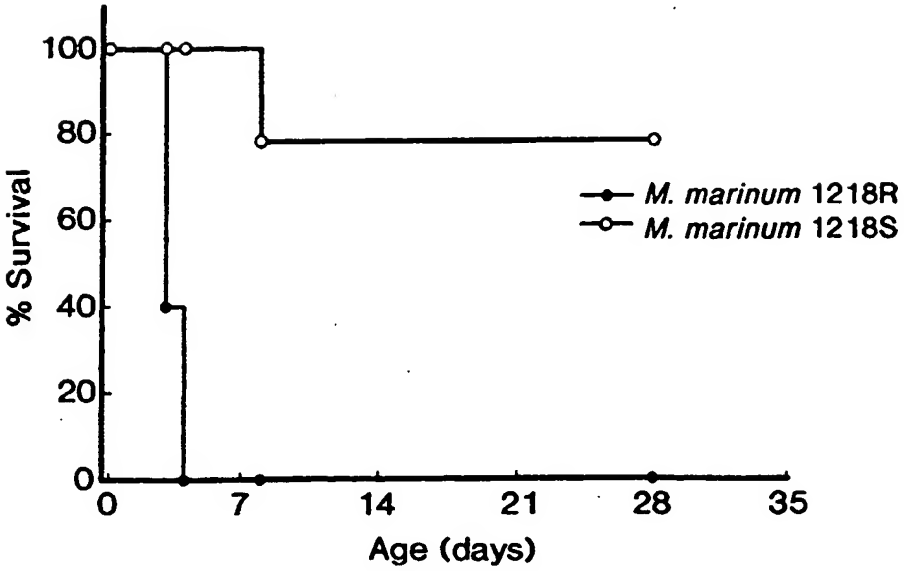
% Survival
100
80
60
40
20
0
0
7
14
21
28
35
Age (days)
M. marinum 1218R
M. marinum 1218S

FIG. 4

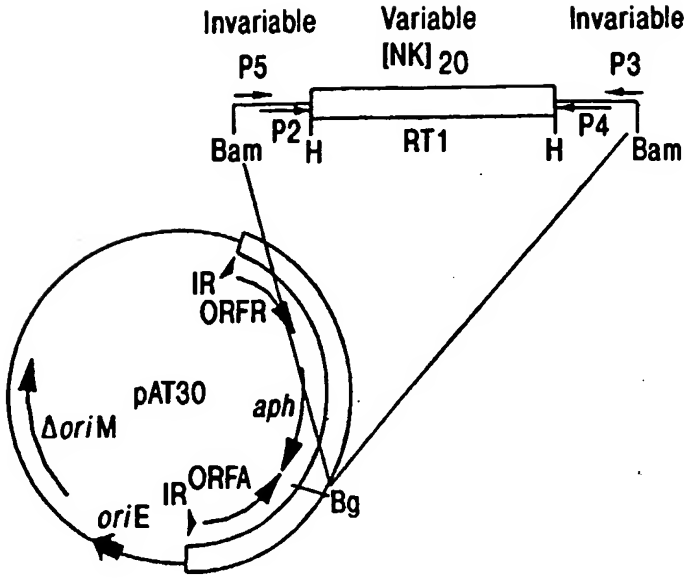
Invariable
Variable
[NK] 20
Invariable
P5
P3
P2
P4
Bam
H
RT1
H
Bam
IR
ORFR
pAT30
aph
ΔoriM
IR
ORFA
oriE
Bg

FIG. 5

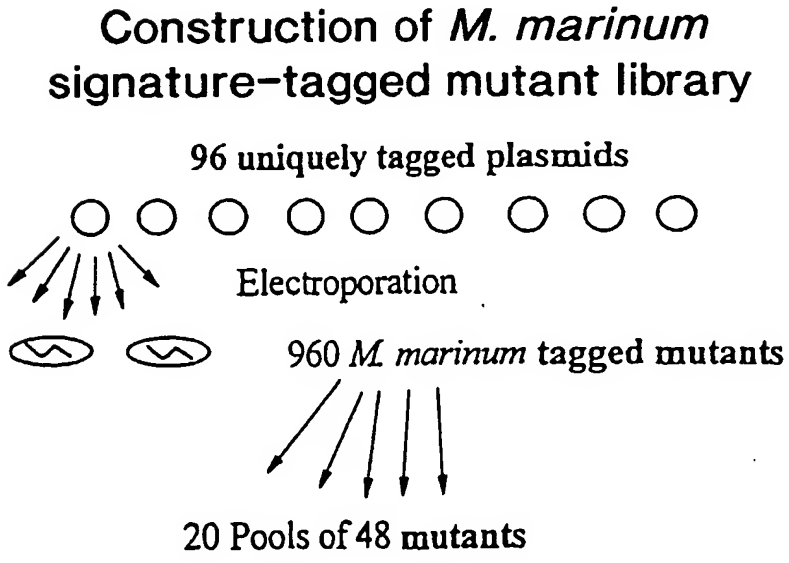
Construction of M. marinum
signature-tagged mutant library
96 uniquely tagged plasmids
Electroporation
960 M. marinum tagged mutants
20 Pools of 48 mutants

FIG. 6

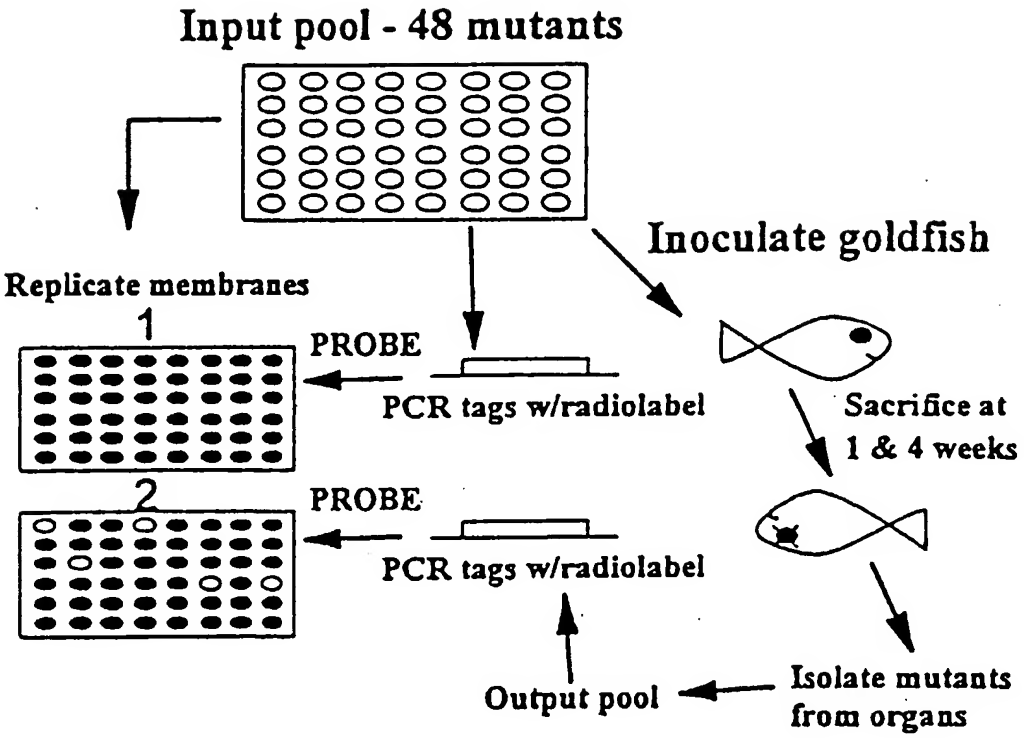
Input pool - 48 mutants
Inoculate goldfish
Replicate membranes
1
PROBE
PCR tags w/radiolabel
Sacrifice at
1 & 4 weeks
2
PROBE
PCR tags w/radiolabel
Output pool
Isolate mutants
from organs

FIG. 7

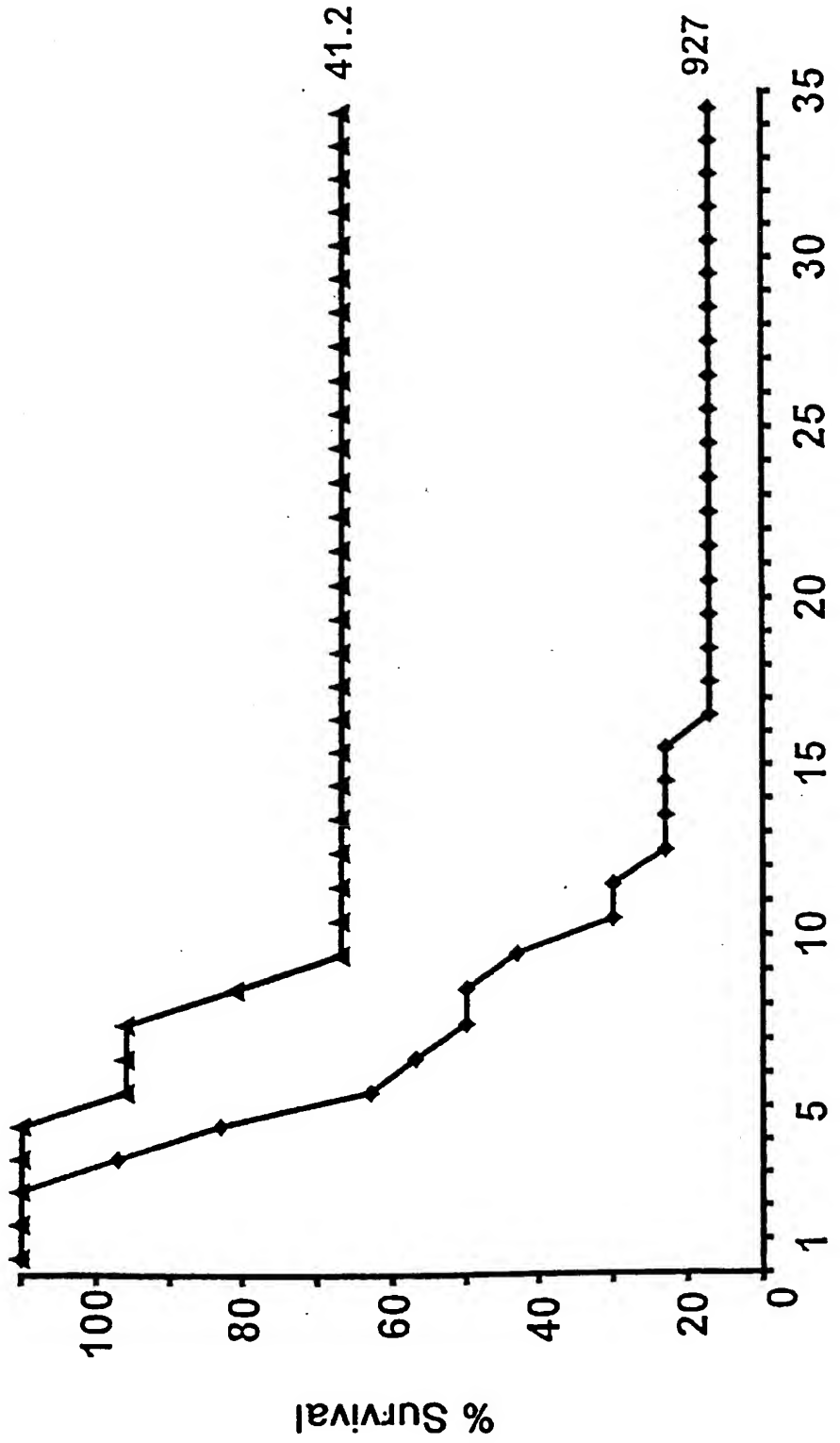

FIG. 8

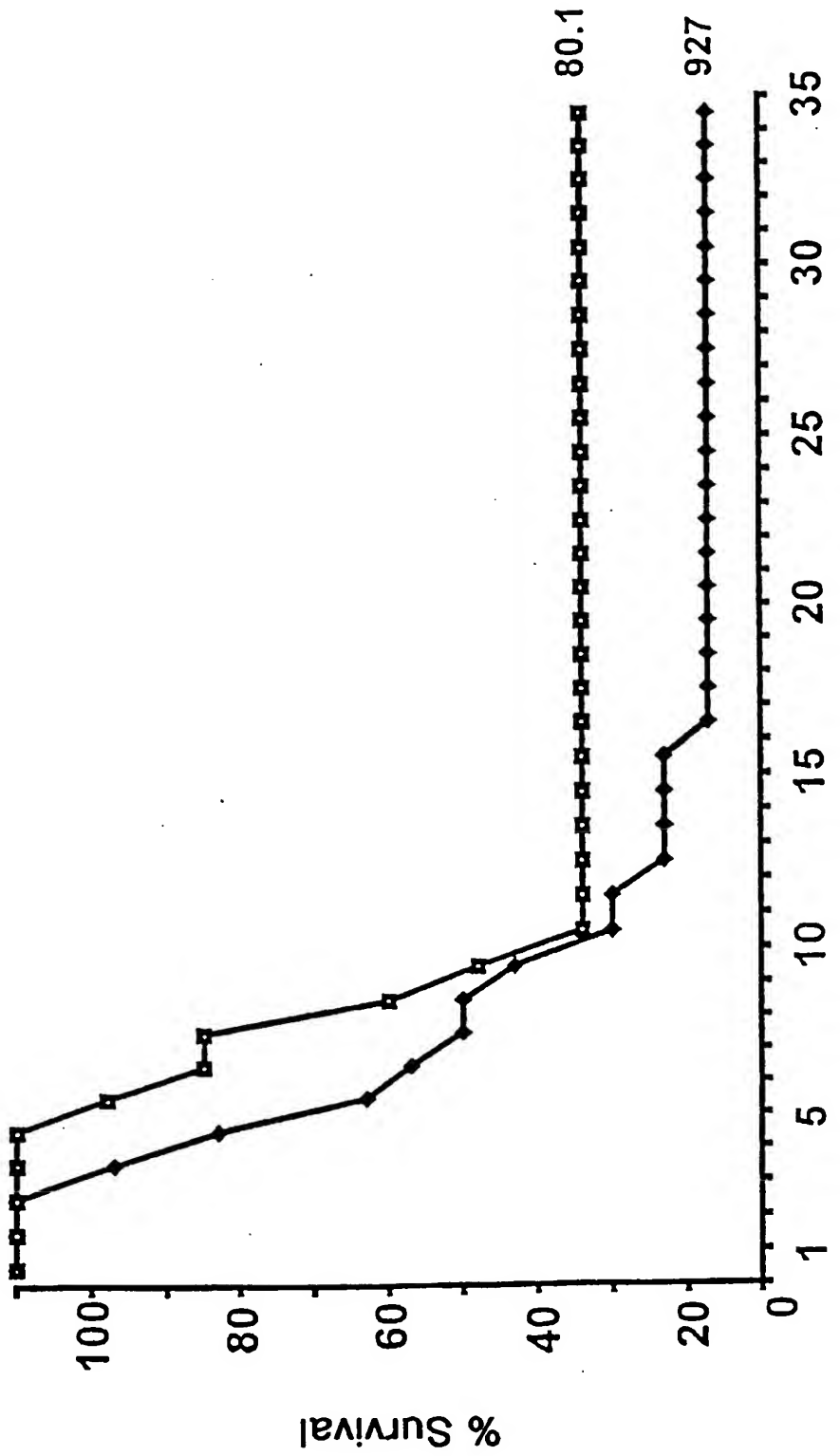

FIG. 9

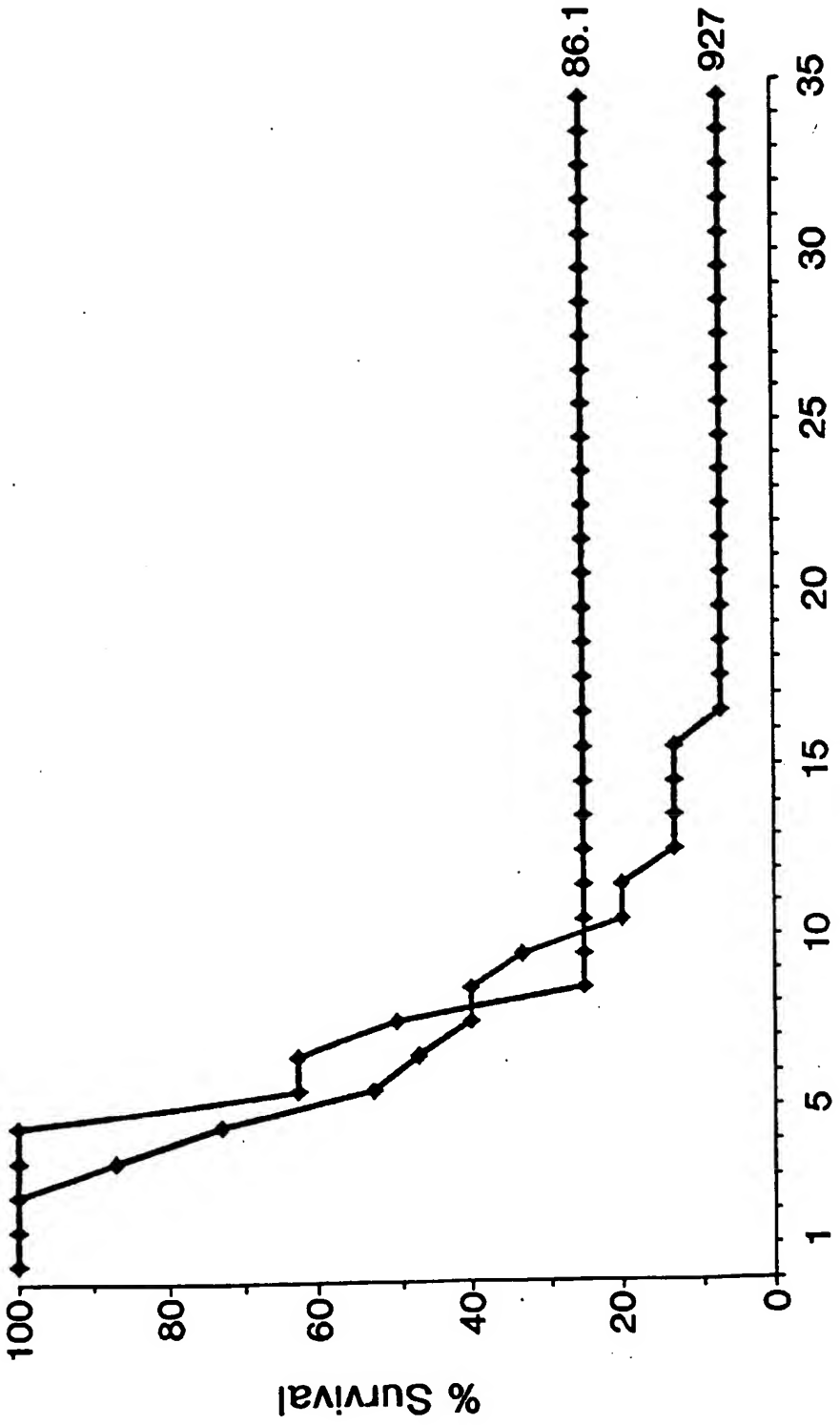

FIG. 10

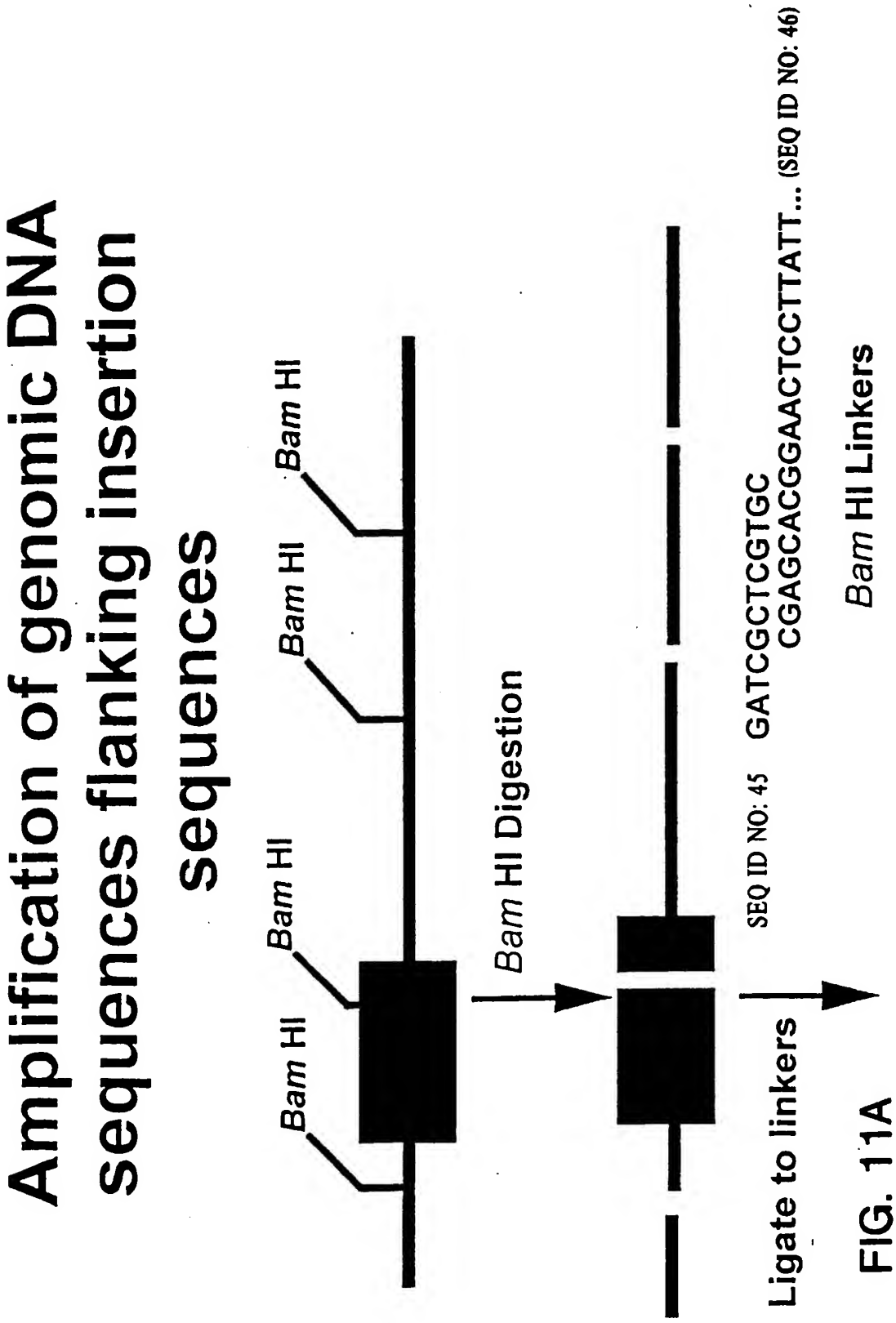
Amplification of genomic DNA sequences flanking insertion sequences
Bam HI
Bam HI
Bam HI
Bam HI Digestion
SEQ ID NO: 45 GATCGCTCGTGC
CGAGCACGGAACTCCTTATT... (SEQ ID NO: 46)
Bam HI Linkers
Ligate to linkers

FIG. 11A

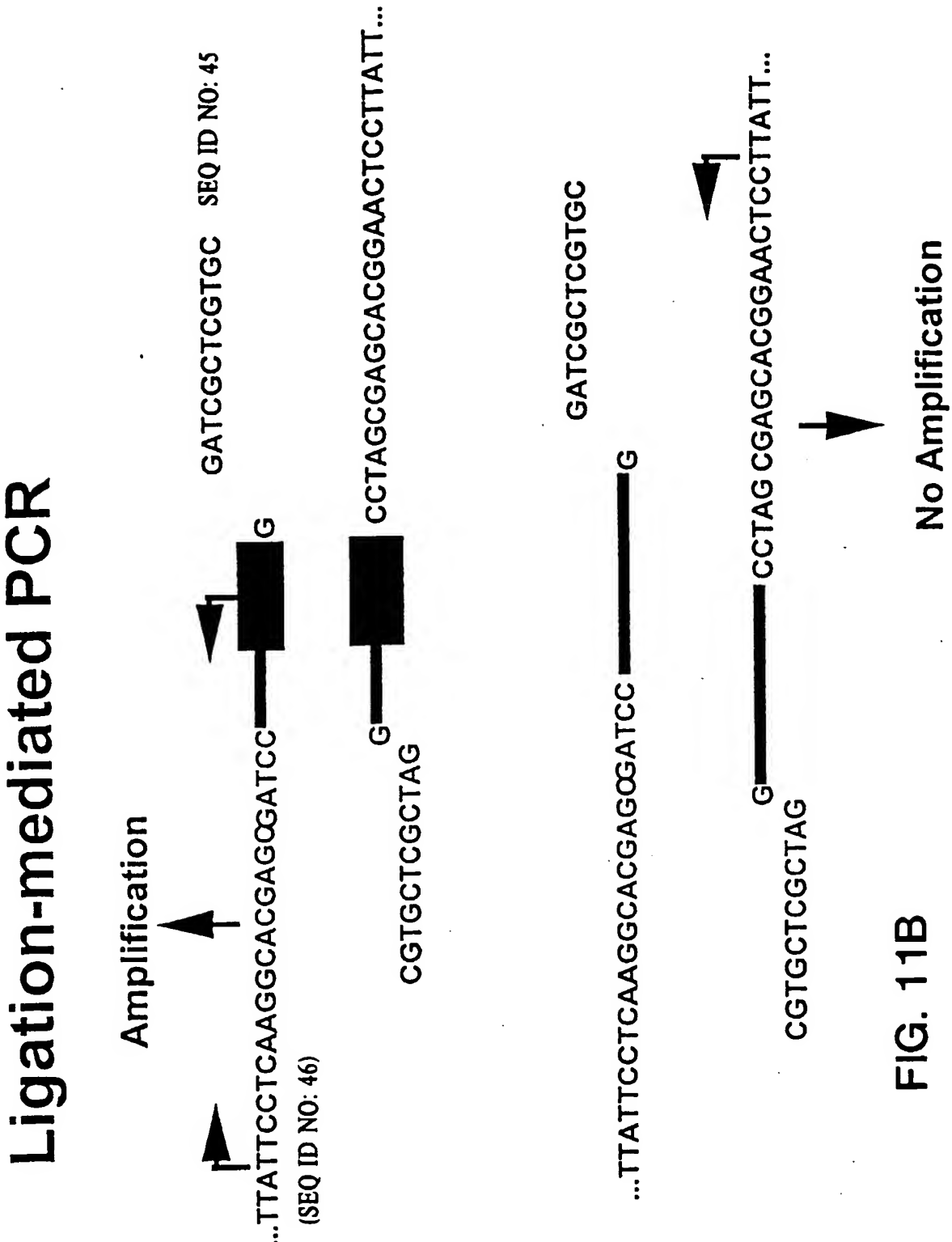
Ligation-mediated PCR
Amplification
GATCGCTCGTGC SEQ ID NO: 45
G
...TTATTCCTCAAGGCACGAGCGGATCC
(SEQ ID NO: 46)
CCTAGCGAGCACGGAACTCCTTATT...
G
CGTGCTCGCTAG
GATCGCTCGTGC
G
...TTATTCCTCAAGGCACGAGCGGATCC
CCTAG CGAGCACGGAACTCCTTATT...
G
CGTGCTCGCTAG
No Amplification

FIG. 11B

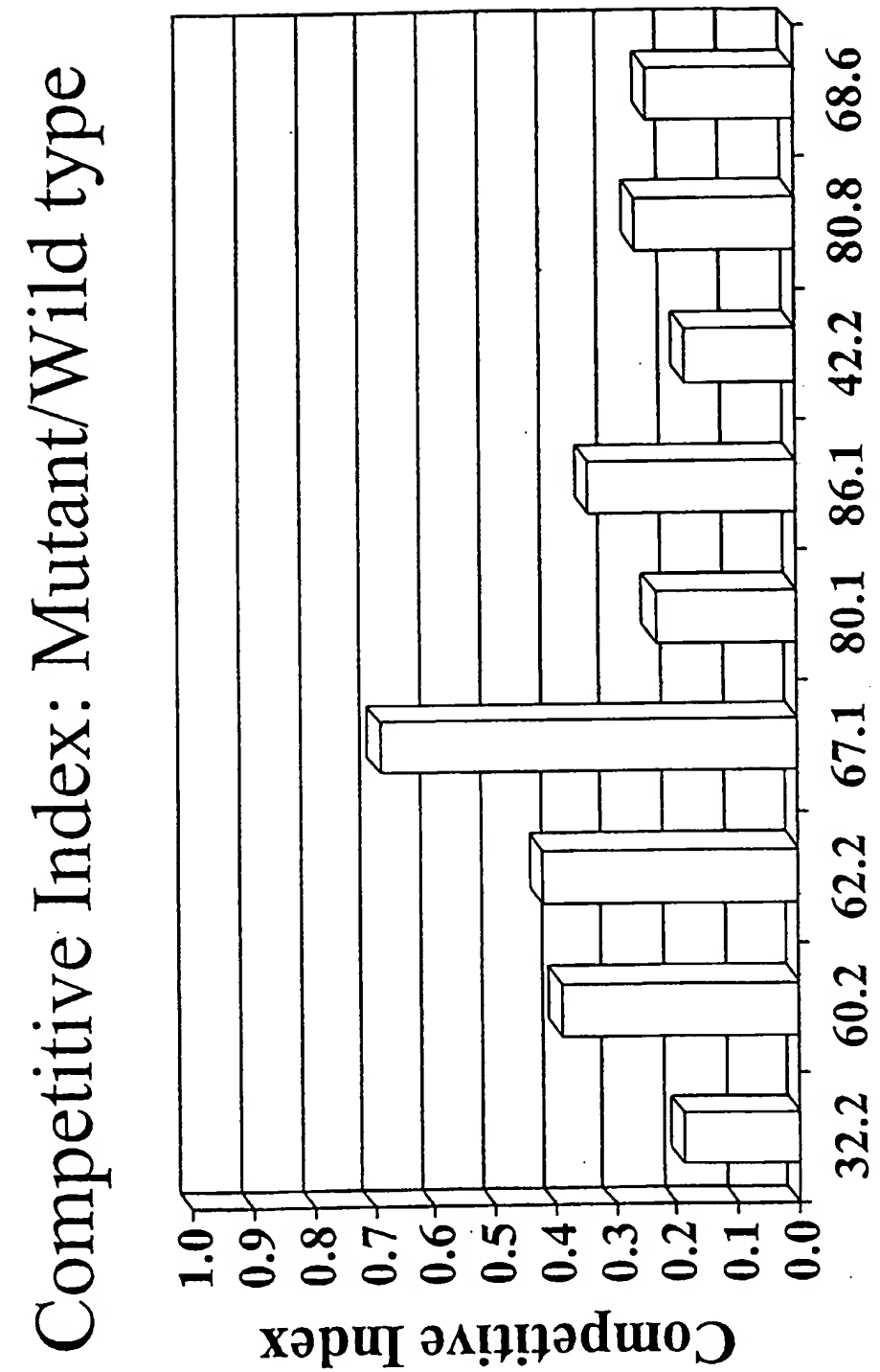
Competitive Index: Mutant/Wild type
Competitive Index
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
32.2
60.2
62.2
67.1
80.1
86.1
42.2
80.8
68.6
Mutant strain

FIG. 12

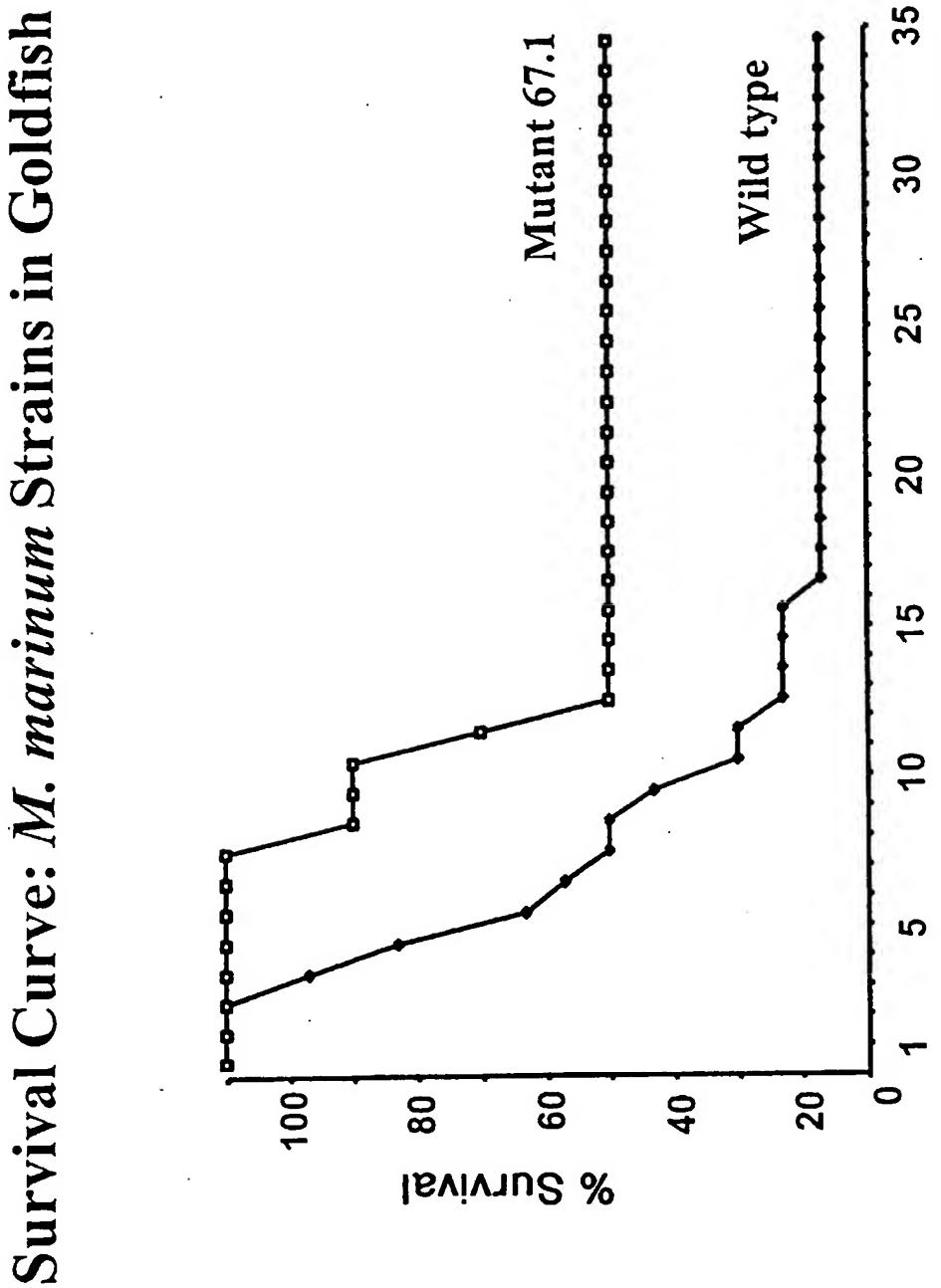
Survival Curve: M. marinum Strains in Goldfish
% Survival
100
80
60
40
20
0
1
5
10
15
20
25
30
35
Age (days)
Mutant 67.1
Wild type

FIG. 13

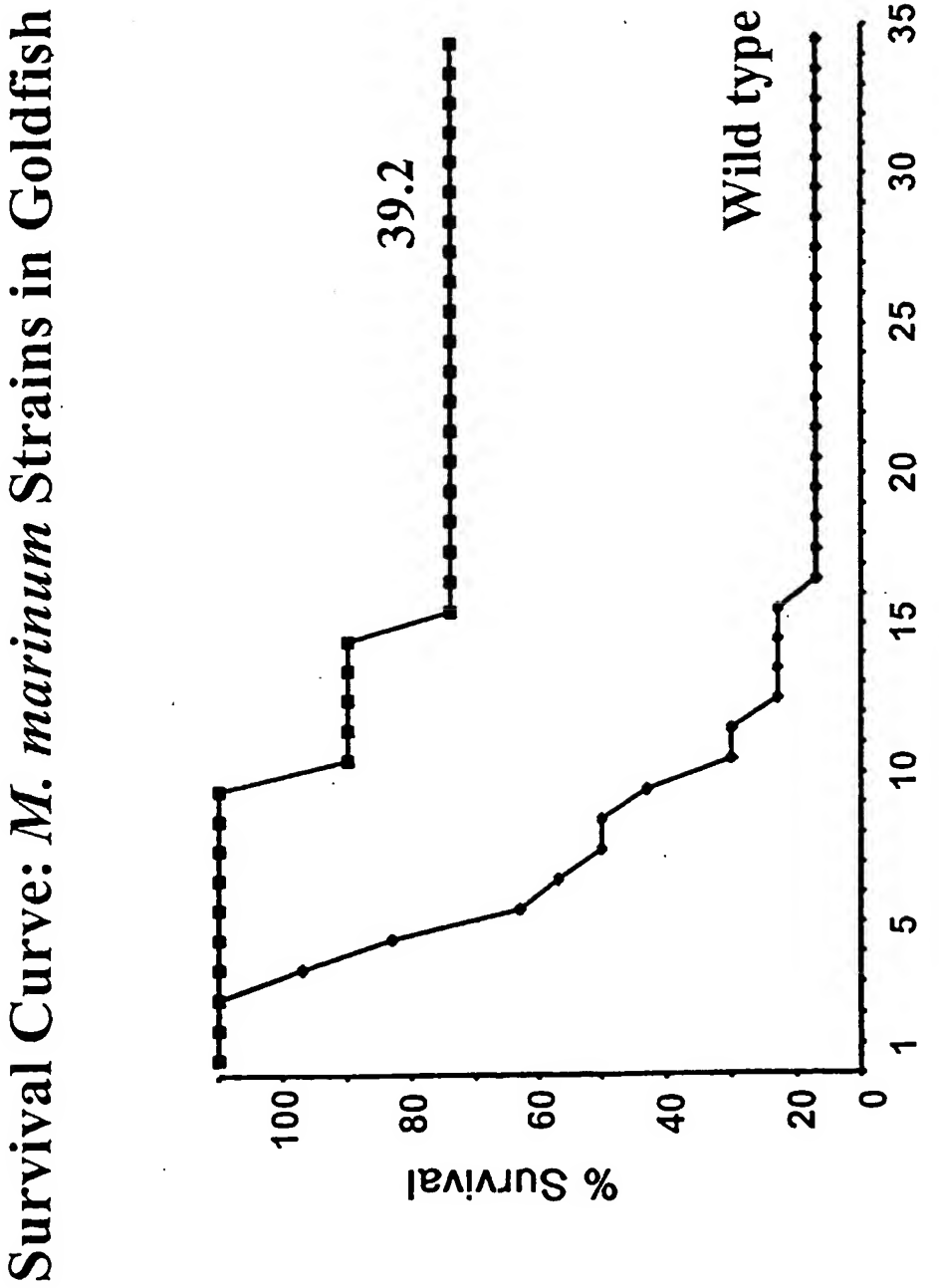
Survival Curve: M. marinum Strains in Goldfish
39.2
Wild type
% Survival
100
80
60
40
20
0
1
5
10
15
20
25
30
35
Age (days)

FIG. 14

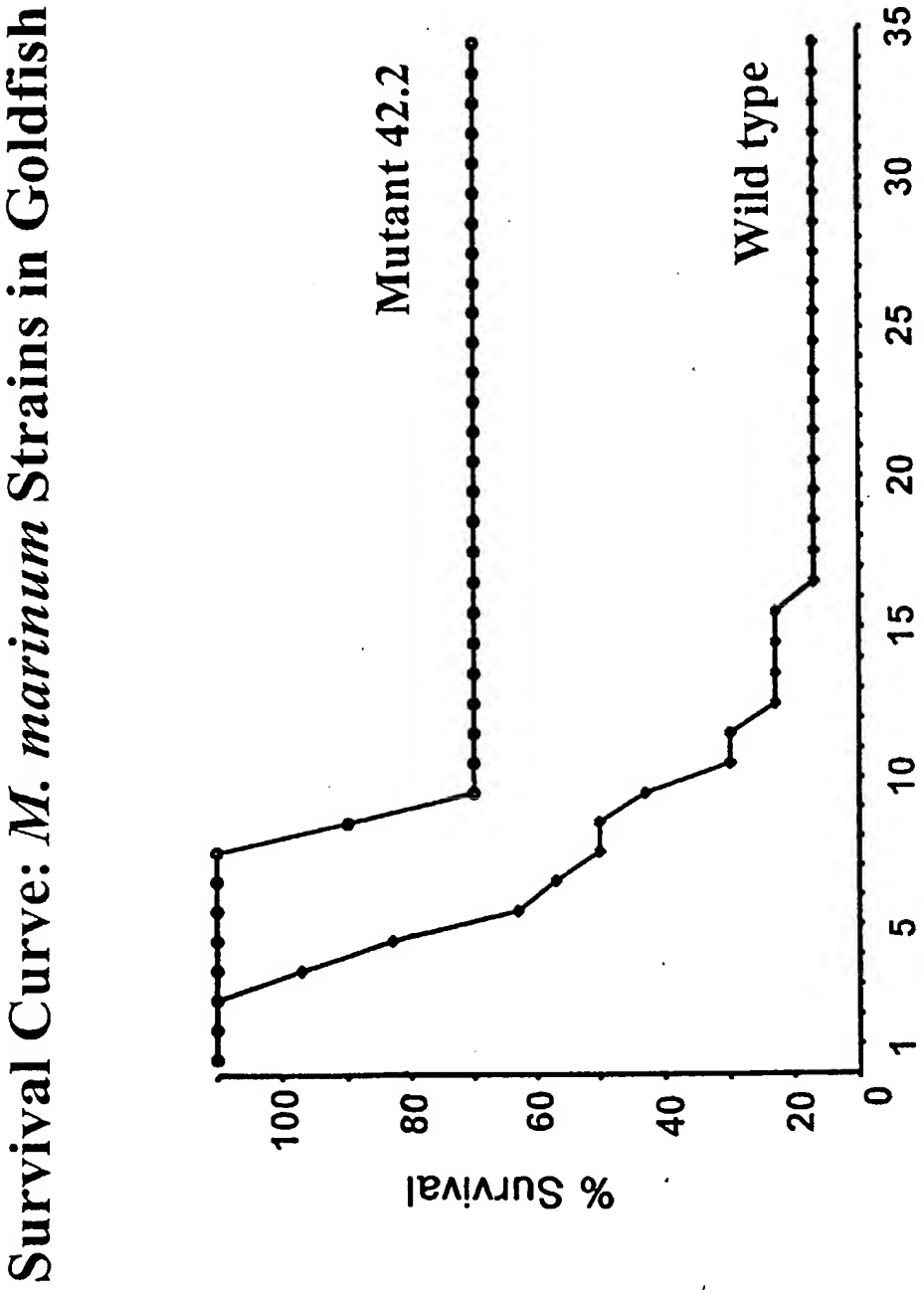
Survival Curve: M. marinum Strains in Goldfish
Mutant 42.2
Wild type
% Survival
100
80
60
40
20
0
1
5
10
15
20
25
30
35
Age (days)

FIG. 15

SEQUENCE LISTING

<110> UNIVERSITY OF MARYLAND
UNITED STATES GOVERMENT, as represented by
Department of Veterans Affairs
TRUCKSIS, Michele

<120> VIRULENCE GENES OF M. MARINUM AND M. TUBERCULOSIS

<130> VET 1 WO

<140>
<141>

<160> 46

<170> PatentIn Ver. 2.1

<210> 1
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 1
ctaggtacct acaacctc 18

<210> 2
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 2
catggtaccc attctaac 18

<210> 3
<211> 89
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: template RT1
oligonucleotide

<220>
<223> "n" represents a, t, c, g, other or unknown

<400> 3
ctaggtacct acaacctcaa gcttnknknk nknknknknk nknknknknk nknknknknk 60
nknkaagctt ggttagaatg ggtaccatg 89

<210> 4
<211> 169
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 41.2

<220>
<223> "n" represents a, t, c, g, other or unknown

<400> 4
cgggccgatc tatgacgagn acgacgggac agatgggtcc ccggatggtc tacaccgaga 60
ccaaactgaa ctcgtcgttc tccttcggcg ggcccaagtg tctggtgaag gtgatccaaa 120
aactgtccgg gttgagcatc aaccggttca tcgccatcga cttcgtcgg 169

<210> 5
<211> 55
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 41.2

<220>
<223> "Xaa" represents any, other or unknown amino acid

<400> 5
Gly Arg Ser Met Thr Xaa Thr Thr Gly Gln Met Gly Pro Arg Met Val
1 5 10 15

Tyr Thr Glu Thr Lys Leu Asn Ser Ser Phe Ser Phe Gly Gly Pro Lys
20 25 30

Cys Leu Val Lys Val Ile Gln Lys Leu Ser Gly Leu Ser Ile Asn Arg
35 40 45

Phe Ile Ala Ile Asp Phe Val
50 55

<210> 6
<211> 382
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 80.1

<220>
<223> "n" represents a, t, c, g, other or unknown

<400> 6
acctcctgaa tgtgtgacat ggccctagaa ccctgcntta gactatttac atacatggct 60
tcacccggcc gcctgtgcca ctcataagac tactggaatg gaccaacaat cgcacagtca 120
tctgaagcag gagtctgtta atcacaggcc ctgaaggaac agtgactgtg cagagaaaga 180

cggcaatgca tcctgttaac taagtggctg gaggagtgcc aggtcattcc aaagaacatc 240
cctgaaatct ggaggagaag gtatagtgag caccccaaaa tttcaactgg agacatcana 300
ccagagtctc tactgagctg ccaagcttgc ggccgcactc gagtaactag ttaacccctt 360
ggggcctcta aacgggtctt ga 382

<210> 7
<211> 121
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 80.1

<220>
<223> "Xaa" represents any, other or unknown amino acid

<400> 7
Pro Pro Glu Cys Val Thr Trp Pro Asn Pro Ala Leu Asp Tyr Leu His
1 5 10 15

Thr Trp Leu His Pro Ala Ala Cys Ala Thr His Lys Thr Thr Gly Met
20 25 30

Asp Gln Gln Ser His Ser His Leu Lys Gln Glu Ser Val Asn His Arg
35 40 45

Pro Arg Asn Ser Asp Cys Ala Glu Lys Asp Gly Asn Ala Ser Cys Leu
50 55 60

Ser Gly Trp Arg Ser Ala Arg Ser Phe Gln Arg Thr Ser Leu Lys Ser
65 70 75 80

Gly Gly Glu Gly Ile Val Ser Thr Pro Lys Phe Gln Leu Glu Thr Ser
85 90 95

Xaa Gln Ser Leu Tyr Ala Ala Lys Leu Ala Ala Ala Leu Glu Leu Val
100 105 110

Asn Pro Leu Gly Pro Leu Asn Gly Ser
115 120

<210> 8
<211> 172
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 86.1

<400> 8
tcatcgctaa ccggttgagc taccgcccgc acagcgtgcc catcatctcc aacctgaccg 60
gctcacttgc cacagtcgag caactcacat cgccccgcta ttgggcacag catgtacggg 120
agccagtgcg gtttcatgac ggcgttaccg gcttgttggc aggcggagaa ca 172

<210> 9

<211> 55
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 86.1

<400> 9
Ala Asn Arg Leu Ser Tyr Arg Pro His Ser Val Pro Ile Ile Ser Asn
1 5 10 15

Leu Thr Gly Ser Leu Ala Thr Val Glu Gln Leu Thr Ser Pro Arg Tyr
20 25 30

Trp Ala Gln His Val Arg Glu Pro Val Arg Phe His Asp Gly Val Thr
35 40 45

Gly Leu Leu Ala Gly Gly Glu
50 55

<210> 10
<211> 228
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 62.2

<400> 10
gatccggtgc cgccttgacc ggccgcgcca ccagtaccgc cgacgccgcc ctggccgccg 60
gcttgtgcgg cttgcgatgg gtcggtgctg tcggtgccgg tgcctccggt gccgccttgg 120
cctccggttc cgccggtgcc gccctggccg ccggcgcctt ggatgccgcc ggtgccggtt 180
ccggctgcac cgcccgttcc gccggttccg cctgcgccgc cggtgcct 228

<210> 11
<211> 225
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 62.2

<220>
<221> CDS
<222> (1)..(225)

<400> 11
ggc acc ggc ggc gca ggc gga acc ggc gga acg ggc ggt gca gcc gga 48
Gly Thr Gly Gly Ala Gly Gly Thr Gly Gly Thr Gly Gly Ala Ala Gly
1 5 10 15

acc ggc acc ggc ggc atc caa ggc gcc ggc ggc cag ggc ggc acc ggc 96
Thr Gly Thr Gly Gly Ile Gln Gly Ala Gly Gly Gln Gly Gly Thr Gly
20 25 30

```
gga acc gga ggc caa ggc ggc acc gga ggc acc ggc acc gac agc acc   144
Gly Thr Gly Gly Gln Gly Gly Thr Gly Gly Thr Gly Thr Asp Ser Thr
         35                  40                  45

gac cca tcg caa gcc gca caa gcc ggc ggc cag ggc ggc gtc ggc ggt   192
Asp Pro Ser Gln Ala Ala Gln Ala Gly Gly Gln Gly Gly Val Gly Gly
     50                  55                  60

act ggt ggc gcg gcc ggt caa ggc ggc acc gga                       225
Thr Gly Gly Ala Ala Gly Gln Gly Gly Thr Gly
 65                  70                  75
```

<210> 12
<211> 75
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 62.2

<400> 12

```
Gly Thr Gly Gly Ala Gly Gly Thr Gly Gly Thr Gly Gly Ala Ala Gly
  1               5                  10                  15

Thr Gly Thr Gly Gly Ile Gln Gly Ala Gly Gly Gln Gly Gly Thr Gly
             20                  25                  30

Gly Thr Gly Gly Gln Gly Gly Thr Gly Gly Thr Gly Thr Asp Ser Thr
         35                  40                  45

Asp Pro Ser Gln Ala Ala Gln Ala Gly Gly Gln Gly Gly Val Gly Gly
     50                  55                  60

Thr Gly Gly Ala Ala Gly Gln Gly Gly Thr Gly
 65                  70                  75
```

<210> 13
<211> 285
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 67.1

<400> 13

```
ggtcgaagac tatcggtatg ctccatagcg ttccgtcggg aagctgcatg ttgtcaaggg 60
tttcgtcgac ctctcggcga cccatgaatc ccgatagtgg cgtgaagaaa ccgtacgaga 120
tgctgatcac ctcgtgggcg gtcgccttcg atatcgggat gcgcaccaat ccctcaatcc 180
ggccggccac gttttccctt tccaccctgt cgacgagtgg gtgtccgtta tggcctaaat 240
aatccatctt gctgcctctt tctgaaatcg aatttattac tatcg                 285
```

<210> 14
<211> 93
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 67.1

<400> 14
Ser Lys Thr Ile Gly Met Leu His Ser Val Pro Ser Gly Ser Cys Met
1 5 10 15

Leu Ser Arg Val Ser Ser Thr Ser Arg Arg Pro Met Asn Pro Asp Ser
20 25 30

Gly Val Lys Lys Pro Tyr Glu Met Leu Ile Thr Ser Trp Ala Val Ala
35 40 45

Phe Asp Ile Gly Met Arg Thr Asn Pro Ser Ile Arg Pro Ala Thr Phe
50 55 60

Ser Leu Ser Thr Leu Ser Thr Ser Gly Cys Pro Leu Trp Pro Lys Ser
65 70 75 80

Ile Leu Leu Pro Leu Ser Glu Ile Glu Phe Ile Thr Ile
85 90

<210> 15
<211> 90
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 67.1

<400> 15
Val Glu Asp Tyr Arg Tyr Ala Pro Arg Ser Val Gly Lys Leu His Val
1 5 10 15

Val Lys Gly Phe Val Asp Leu Ser Ala Thr His Glu Ser Arg Trp Arg
20 25 30

Glu Glu Thr Val Arg Asp Ala Asp His Leu Val Gly Gly Arg Leu Arg
35 40 45

Tyr Arg Asp Ala His Gln Ser Leu Asn Pro Ala Gly His Val Phe Pro
50 55 60

Phe His Pro Val Asp Glu Trp Val Ser Val Met Ala Ile Ile His Leu
65 70 75 80

Ala Ala Ser Phe Asn Arg Ile Tyr Tyr Tyr
85 90

<210> 16
<211> 92
<212> PRT
<213> Mycobacterium marinum

<220>
<223> 67.1

<400> 16
Gly Arg Arg Leu Ser Val Cys Ser Ile Ala Phe Arg Arg Glu Ala Ala
1 5 10 15

Cys Cys Gln Gly Phe Arg Arg Pro Leu Gly Asp Pro Ile Pro Ile Val
20 25 30

Ala Arg Asn Arg Thr Arg Cys Ser Pro Arg Gly Arg Ser Pro Ser Ile
35 40 45

Ser Gly Cys Ala Pro Ile Pro Gln Ser Gly Arg Pro Arg Phe Pro Phe
50 55 60

Pro Pro Cys Arg Arg Val Gly Val Arg Tyr Gly Leu Asn Asn Pro Ser
65 70 75 80

Cys Cys Leu Phe Leu Lys Ser Asn Leu Leu Leu Ser
85 90

<210> 17
<211> 285
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 67.1

<400> 17
cgatagtaat aaattcgatt tcagaaagag gcagcaagat ggattattta ggccataacg 60
gacacccact cgtcgacagg gtggaaaggg aaaacgtggc cggccggatt gagggattgg 120
tgcgcatccc gatatcgaag gcgaccgccc acgaggtgat cagcatctcg tacggtttct 180
tcacgccact atcgggattc atgggtcgcc gagaggtcga cgaaaccctt gacaacatgc 240
agcttcccga cggaacgcta tggagcatac cgatagtctt cgacc 285

<210> 18
<211> 89
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 67.1

<400> 18
Arg Ile Arg Phe Gln Lys Glu Ala Ala Arg Trp Ile Ile Ala Ile Thr
1 5 10 15

Asp Thr His Ser Ser Thr Gly Trp Lys Gly Lys Thr Trp Pro Ala Gly
20 25 30

Leu Arg Asp Trp Cys Ala Ser Arg Tyr Arg Arg Arg Pro Pro Thr Arg
35 40 45

Ser Ala Ser Arg Thr Val Ser Ser Arg His Tyr Arg Asp Ser Trp Val
50 55 60

Ala Glu Arg Ser Thr Lys Pro Leu Thr Thr Cys Ser Phe Pro Thr Glu
65 70 75 80

Arg Tyr Gly Ala Tyr Arg Ser Ser Thr
85

<210> 19
<211> 91
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 67.1

<400> 19
Asp Ser Asn Lys Phe Asp Phe Arg Lys Arg Gln Gln Asp Gly Leu Phe
1 5 10 15

Arg Pro Arg Thr Pro Thr Arg Arg Gln Gly Gly Lys Gly Lys Arg Gly
20 25 30

Arg Pro Asp Gly Ile Gly Ala His Pro Asp Ile Glu Gly Asp Arg Pro
35 40 45

Arg Gly Asp Gln His Leu Val Arg Phe Leu His Ala Thr Ile Gly Ile
50 55 60

His Gly Ser Pro Arg Gly Arg Arg Asn Pro Gln His Ala Ala Ser Arg
65 70 75 80

Arg Asn Ala Met Glu His Thr Asp Ser Leu Arg
85 90

<210> 20
<211> 94
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 67.1

<400> 20
Ile Val Ile Asn Ser Ile Ser Glu Arg Gly Ser Lys Met Asp Tyr Leu
1 5 10 15

Gly His Asn Gly His Pro Leu Val Asp Arg Val Glu Arg Glu Asn Val
20 25 30

Ala Gly Arg Ile Glu Gly Leu Val Arg Ile Pro Ile Ser Lys Ala Thr
35 40 45

Ala His Glu Val Ile Ser Ile Ser Tyr Gly Phe Phe Thr Pro Leu Ser
50 55 60

Gly Phe Met Gly Arg Arg Glu Val Asp Glu Thr Leu Asp Asn Met Gln
65 70 75 80

Leu Pro Asp Gly Thr Leu Trp Ser Ile Pro Ile Val Phe Asp
85 90

<210> 21
<211> 167
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 80.8

<400> 21
ccaattagct gattattcct cgggcgtgct caacgccaag gactacatat caggttactt 60
ccactaaaat tcgcgggccc cgatcggcga cattactcga cggttttcgg gggaatctca 120
gcggtgatgg cattcttgag ggcgacgtag cgtttggcgt cgggatc 167

<210> 22
<211> 53
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 80.8

<400> 22
Asp Pro Asp Ala Lys Arg Tyr Val Ala Leu Lys Asn Ala Ile Thr Ala
1 5 10 15

Glu Ile Pro Pro Lys Thr Val Glu Cys Arg Arg Ser Gly Pro Ala Asn
20 25 30

Phe Ser Gly Ser Asn Leu Ile Cys Ser Pro Trp Arg Ala Arg Pro Arg
35 40 45

Asn Asn Gln Leu Ile
50

<210> 23
<211> 144
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 39.

<400> 23
gatccgctgg acggcaccaa agaattcatc aagggcagcg atgagttcac cgtcaacatc 60
gccctggtcg agaaccagga acccattctc ggggcaatct acggtccagc gaagcaactt 120
ctgcactacg cggccaaagg ggct 144

<210> 24
<211> 46
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 39.2

<400> 24

Leu Asp Gly Thr Lys Glu Phe Ile Lys Gly Ser Asp Glu Phe Thr Val
1 5 10 15

Asn Ile Ala Leu Val Glu Asn Gln Glu Pro Ile Leu Gly Ala Ile Tyr
20 25 30

Gly Pro Ala Lys Gln Leu Leu His Tyr Ala Ala Lys Gly Ala
35 40 45

<210> 25
<211> 381
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 114.7

<220>
<223> "n" represents a, t, c, g, other or unknown

<400> 25
agccgtattt cgccattgag agttggggtc ttgagatcgg cactggaagg ggacagcgtg 60
ctattgcctc ttggtccgcc cttgccacct gatgctgtgg cggctaaacg gggtgagtcg 120
gggctgctct gcggcttgtc ggttccgctc agctggggta cggccgttcc gccggatgac 180
tacnaccatt gggcaccgga gcctgaagaa ggcgccgagg ccgtggtcga agaaaacgtg 240
gatgcggcag ctgccggtac cgacgagtgg gacgagtggg cggaatggag ggagtgggag 300
gcagcaaatg cccgaacctc attttcgaga tgccccgtac cagcagccgt gatacccgaa 360
ctcgccggcg gccggttgag a 381

<210> 26
<211> 122
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 114.7

<220>
<223> "Xaa" represents any, other or unknown amino acid

<400> 26
Leu Arg Val Gly Val Leu Arg Ser Ala Leu Glu Gly Asp Ser Val Leu
1 5 10 15

Leu Pro Leu Gly Pro Pro Leu Pro Pro Asp Ala Val Ala Ala Lys Arg
20 25 30

Gly Glu Ser Gly Leu Leu Cys Gly Leu Ser Val Pro Leu Ser Trp Gly
35 40 45

Thr Ala Val Pro Pro Asp Asp Tyr Xaa His Trp Ala Pro Glu Pro Glu
50 55 60

Glu Gly Ala Glu Ala Val Val Glu Glu Asn Val Asp Ala Ala Ala Ala
65 70 75 80

Gly Thr Asp Glu Trp Asp Glu Trp Ala Glu Trp Arg Glu Trp Glu Ala
85 90 95

Ala Asn Ala Arg Thr Ser Phe Ser Arg Cys Pro Val Pro Ala Ala Val
100 105 110

Ile Pro Glu Leu Ala Gly Gly Arg Leu Arg
115 120

<210> 27
<211> 98
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 32.2

<220>
<223> "n" represents a, t, c, g, other or unknown

<400> 27
tccanncaga ggngcacgta gancgtagga cggaangcgg ngngatcgnc aatacggctg 60
gcnctgcnag aactgntcga gggcctgcng ctggggcc 98

<210> 28
<211> 32
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 32.2

<220>
<223> "Xaa" represents any, other or unknown amino acid

<400> 28
Ala Pro Ala Ala Gly Pro Arg Xaa Val Leu Ala Xaa Pro Ala Val Leu
1 5 10 15

Xaa Ile Xaa Pro Xaa Ser Val Leu Arg Ser Thr Cys Xaa Ser Xaa Trp
20 25 30

<210> 29
<211> 62
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 42.2

<220>
<223> "n" represents a, t, c, g, other or unknown

<400> 29
tttgcaatcc acctgtacgc ggaactnttn annnccgttt tgccttgncg aataagctag 60

ct 62

<210> 30
<211> 19
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 42.2

<220>
<223> "Xaa" represents any, other or unknown amino acid

<400> 30
Ser Leu Ile Arg Gln Gly Lys Thr Xaa Xaa Xaa Ser Ser Ala Tyr Arg
1 5 10 15

Trp Ile Ala

<210> 31
<211> 74
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 60.2

<220>
<223> "n" represents a, t, c, g, other or unknown

<400> 31
ccanacctat ctgtttncag nttnagacna cggnatctca cgcgnttggg cccngccacc 60
aaacgccgcg tnga 74

<210> 32
<211> 24
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 60.2

<220>
<223> "Xaa" represents any, other or unknown amino acid

<400> 32
Xaa Pro Ile Cys Xaa Gln Xaa Xaa Thr Thr Xaa Ser His Ala Xaa Gly
1 5 10 15

Pro Xaa His Gln Thr Pro Arg Xaa
20

<210> 33
<211> 24
<212> PRT

<213> Mycobacterium marinum

<220>
<223> Mutant 60.2

<220>
<223> "Xaa" represents any, other or unknown amino acid

<400> 33
Xaa Thr Tyr Leu Phe Xaa Xaa Xaa Asp Xaa Gly Ile Ser Arg Xaa Trp
1 5 10 15

Ala Xaa Pro Pro Asn Ala Ala Xaa
20

<210> 34
<211> 24
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 60.2

<220>
<223> "Xaa" represents any, other or unknown amino acid

<400> 34
Pro Xaa Leu Ser Val Xaa Xaa Xaa Arg Xaa Arg Xaa Leu Thr Arg Leu
1 5 10 15

Gly Pro Ala Thr Lys Arg Arg Val
20

<210> 35
<211> 74
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 60.2

<220>
<223> "n" represents a, t, c, g, other or unknown

<400> 35
tcnacgcggc gtttggtggc ngggcccaan cgcgtgagat nccgtngtct naanctgnaa 60
acagataggt ntgg 74

<210> 36
<211> 24
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 60.2

<220>

<223> "Xaa" represents any, other or unknown amino acid

<400> 36
Ser Thr Arg Arg Leu Val Ala Gly Pro Xaa Arg Val Arg Xaa Arg Xaa
1 5 10 15

Leu Xaa Leu Xaa Thr Asp Arg Xaa
20

<210> 37
<211> 23
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 60.2

<220>
<223> "Xaa" represents any, other or unknown amino acid

<400> 37
Xaa Arg Gly Val Trp Trp Xaa Gly Pro Xaa Ala Asp Xaa Val Val Xaa
1 5 10 15

Xaa Xaa Lys Gln Ile Gly Xaa
20

<210> 38
<211> 23
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 60.2

<220>
<223> "Xaa" represents any, other or unknown amino acid

<400> 38
Xaa Ala Ala Phe Gly Gly Xaa Ala Gln Xaa Arg Glu Xaa Pro Xaa Ser
1 5 10 15

Xaa Xaa Xaa Asn Arg Val Trp
20

<210> 39
<211> 247
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 68.6

<400> 39
aaatcatcat ctatcgttac ccggggcaag ccaagcacct cagcaaaaat tctgcagagc 60
atttcctctt gcggagttcg cggcatacgg ccaatcgccg catgatgatc gggcacaggc 120
agcgctttac gatccaccct cttattcgga gttaacggca tggtctcaag tcttacgatg 180
acagacggca ccatatattc ggccagtttc agggaggcgt agcgccgcag ttctgctgta 240

tctatca 247

<210> 40
<211> 81
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 68.6

<400> 40
Ile Asp Thr Ala Glu Leu Arg Arg Tyr Ala Ser Leu Lys Leu Ala Glu
1 5 10 15

Tyr Met Val Pro Ser Val Ile Val Arg Leu Glu Thr Met Pro Leu Thr
20 25 30

Pro Asn Lys Lys Val Asp Arg Lys Ala Leu Pro Val Pro Asp His His
35 40 45

Ala Ala Ile Gly Arg Met Pro Arg Thr Pro Gln Glu Glu Met Leu Cys
50 55 60

Arg Ile Phe Ala Glu Val Leu Gly Leu Pro Arg Val Thr Ile Asp Asp
65 70 75 80

Asp

<210> 41
<211> 164
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 95.3

<400> 41
gattagctta ttcctcaagg cacgagcgat tagcttattc ctcaaggcac gagcgactag 60
cttattcctc aaggcacgag cttcgcactt gacggtgtag agctcaatag cttattcctc 120
aaggcacgag ctcgacttcg cacttgacgg tgtagagctc aaag 164

<210> 42
<211> 50
<212> PRT
<213> Mycobacterium marinum

<220>
<223> Mutant 95.3

<400> 42
Asp Leu Ile Pro Gln Gly Thr Ser Asp Leu Ile Pro Gln Gly Thr Ser
1 5 10 15

Asp Leu Ile Pro Gln Gly Thr Ser Phe Ala Leu Asp Gly Val Glu Leu
20 25 30

```
Asn Ser Leu Phe Leu Lys Ala Arg Ala Arg Leu Arg Thr Arg Cys Arg
         35                  40                  45

Ala Gln
     50
```

<210> 43
<211> 138
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 39.2

<220>
<221> CDS
<222> (1)..(138)

<400> 43

```
ctg gac ggc acc aaa gaa ttc atc aag ggc agc gat gag ttc acc gtc   48
Leu Asp Gly Thr Lys Glu Phe Ile Lys Gly Ser Asp Glu Phe Thr Val
  1               5                  10                  15

aac atc gcc ctg gtc gag aac cag gaa ccc att ctc ggg gca atc tac   96
Asn Ile Ala Leu Val Glu Asn Gln Glu Pro Ile Leu Gly Ala Ile Tyr
             20                  25                  30

ggt cca gcg aag caa ctt ctg cac tac gcg gcc aaa ggg gct           138
Gly Pro Ala Lys Gln Leu Leu His Tyr Ala Ala Lys Gly Ala
         35                  40                  45
```

<210> 44
<211> 366
<212> DNA
<213> Mycobacterium marinum

<220>
<223> Mutant 114.7

<220>
<223> "n" represents a, t, c, g, other or unknown

<220>
<221> CDS
<222> (1)..(366)

<400> 44

```
ttg aga gtt ggg gtc ttg aga tcg gca ctg gaa ggg gac agc gtg cta   48
Leu Arg Val Gly Val Leu Arg Ser Ala Leu Glu Gly Asp Ser Val Leu
  1               5                  10                  15

ttg cct ctt ggt ccg ccc ttg cca cct gat gct gtg gcg gct aaa cgg   96
Leu Pro Leu Gly Pro Pro Leu Pro Pro Asp Ala Val Ala Ala Lys Arg
             20                  25                  30
```

```
ggt gag tcg ggg ctg ctc tgc ggc ttg tcg gtt ccg ctc agc tgg ggt   144
Gly Glu Ser Gly Leu Leu Cys Gly Leu Ser Val Pro Leu Ser Trp Gly
         35                  40                  45

acg gcc gtt ccg ccg gat gac tac nac cat tgg gca ccg gag cct gaa   192
Thr Ala Val Pro Pro Asp Asp Tyr Xaa His Trp Ala Pro Glu Pro Glu
     50                  55                  60

gaa ggc gcc gag gcc gtg gtc gaa gaa aac gtg gat gcg gca gct gcc   240
Glu Gly Ala Glu Ala Val Val Glu Glu Asn Val Asp Ala Ala Ala Ala
 65                  70                  75                  80

ggt acc gac gag tgg gac gag tgg gcg gaa tgg agg gag tgg gag gca   288
Gly Thr Asp Glu Trp Asp Glu Trp Ala Glu Trp Arg Glu Trp Glu Ala
                 85                  90                  95

gca aat gcc cga acc tca ttt tcg aga tgc ccc gta cca gca gcc gtg   336
Ala Asn Ala Arg Thr Ser Phe Ser Arg Cys Pro Val Pro Ala Ala Val
            100                 105                 110

ata ccc gaa ctc gcc ggc ggc cgg ttg aga                          366
Ile Pro Glu Leu Ala Gly Gly Arg Leu Arg
        115                 120
```

<210> 45
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Oligonucleotide

<400> 45
gatcgctcgt gc 12

<210> 46
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Oligonucleotide

<400> 46
ttattcctca aggcacgagc gatcc 25

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 March 2001 (22.03.2001)

PCT

(10) International Publication Number
WO 01/19993 A3

(51) International Patent Classification[7]: C12N 15/13, C07K 14/35, A61K 38/00, 39/00, A61P 31/04, C12N 9/00

(21) International Application Number: PCT/US00/25512

(22) International Filing Date:
18 September 2000 (18.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/154,322 17 September 1999 (17.09.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/154,322 (CIP)
Filed on 17 September 1999 (17.09.1999)

(71) Applicants *(for all designated States except US)*: **UNIVERSITY OF MARYLAND, BALTIMORE** [US/US]; 520 West Lombard Avenue, Baltimore, MD 21201 (US). **UNITED STATES GOVERNMENT, as represented by DEPARTMENT OF VETERANS AFFAIRS** [US/US]; 810 Vermont Avenue, N.W., Washington, DC 20420 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **TRUCKSIS, Michele** [US/US]; 5072 Jericho Road, Columbia, MD 21044 (US).

(74) Agents: **ZELANO, Anthony, J.** et al.; Millen, White, Zelano & Branigan, P.C., Arlington Courthouse Plaza 1, 2200 Clarendon Boulevard, Suite 1400, Arlington, VA 22201 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
22 November 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/19993 A3

(54) Title: VIRULENCE GENES OF *M. MARINUM* AND *M. TUBERCULOSIS*

(57) Abstract: Methods for identifying, isolating and mutagenizing virulence genes of mycobacteria, *e.g.*, *M. marinum* and *M. tuberculosis*, are described. Also described are isolated virulence genes and fragments of them, isolated gene products and fragments of them, avirulent bacteria in which one or more virulence genes are mutagenized, attenuated vaccines containing such mutant bacteria, and methods to elicit an immune response in a host, using such mutant bacteria.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/25512

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/13 C07K14/35 A61K38/00 A61K39/00 A61P31/04 C12N9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K A61K A61P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, PAJ, WPI Data, EMBL, BIOSIS, EMBASE, MEDLINE, CAB Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	A. TALAAT ET AL.: "Use of signature-tagged mutagenesis to identify Mycobacterium marinum genes required for in vivo survival in the goldfish model of Mycobacterial infection." 99TH GENERAL MEETINF OF THE AMERICAL SOCIETY FOR MICROBIOLOGY, ABSTRACTS IN MICROBIAL PATHOGENESIS/GENERAL MEDICAL MICROBIOLOGY, vol. 99, no. 29/B/D, Abstract B/D-15, 30 May 1999 (1999-05-30) - 3 June 1999 (1999-06-03), pages 31-32, XP000978661 Chicago, Illinois, USA	1-5,50
Y	abstract	28-42, 51-76
	--- -/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 March 2001	11 06 2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hix, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/25512

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	COLE S T ET AL: "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence" NATURE,MACMILLAN JOURNALS LTD. LONDON,GB, vol. 393, no. 6685, 11 June 1998 (1998-06-11), pages 537-544, XP002154587 ISSN: 0028-0836 the whole document	28,31, 35,51, 54,58
X	WO 99 09186 A (PORTNOI DENIS ;GUIGUENO AGNES (FR); LIM ENG MONG (FR); GICQUEL BRI) 25 February 1999 (1999-02-25) the whole document	28,51
X	DATABASE EMBL [Online] Accession number: AI592454, 26 April 1999 (1999-04-26) M. MARRA ET AL.: "The WashU-NCI mouse EST Project 1999" XP002161752 sequence data abstract	29,52
Y	TALAAT A.M. ET AL: "Goldfish, Carassius auratus, a novel animal model for the study of Mycobacterium marinum pathogenesis." INFECTION AND IMMUNITY, (1998) 66/6 (2938-2942)., XP000979750 the whole document	1-5, 28-42, 50-76
Y	A. TALAAT ET AL.: "Goldfish, Carassius auratus a fish model for mycobacterial disease." 96TH GENERAL MEETING OF THE AMERICAL SOCIETY FOR MICROBIOLOGY, vol. 96, no. U-177, 19 - 23 May 1996, page 132 XP000978662 New Orleans, Louisiane, USA abstract	1-5, 28-42, 50-76
P,X	TALAAT A M ET AL: "Transformation and transposition of the genome of Mycobacterium marinum." AMERICAN JOURNAL OF VETERINARY RESEARCH, (2000 FEB) 61 (2) 125-8., XP000981305 the whole document	1-5, 28-42, 50-76

-/--

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/25512

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,Y	L. RAMAKRISHNAN ET AL.: "Granuloma-specific expression of Mycobacterium virulence proteins from the Glycine-rich PE-PGRS family." SCIENCE, vol. 288, 26 May 2000 (2000-05-26), pages 1436-1439, XP000979218 the whole document	1-5, 28-42, 50-76
A	RINDI L. ET AL: "Search for genes potentially involved in Mycobacterium tuberculosis virulence by mRNA differential display." BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, (29 APR 1999) 258/1 (94-101)., XP000979851 the whole document	
A	A. KIRAN KINGER ET AL.: "Identification and cloning of genes differentially expressed in the virulent strain of Mycobacterium tuberculosis." GENE, vol. 131, 1993, pages 113-117, XP000978684 the whole document	
A	K. SRIVASTAVA ET AL.: "Immunogenic behaviour of Mycobacterium marinum (SATO) in mice." THE INDIAN JOURNAL OF MEDICAL RESEARCH, vol. 84, 1986, pages 485-491, XP000979839 the whole document	
A	T. TONJUM ET AL.: "Differentiation of Mycobacterium ulcerans, M. marinum and M. haemophilum: Mapping of thier relationships to M. tuberculosis by fatty actid profile analysis, DNA-DNA Hybridization and 16S rRNA gene sequence analysis." JOURNAL OF CLINICAL MICROBIOLOGY, vol. 36, no. 4, April 1998 (1998-04), pages 918-925, XP000979751 cited in the application the whole document	

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 00/25512

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 46 and 47 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-5, 28-42, 46, 50-76

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-5, 28-42, 46, 50-76

Method for identifying a virulence gene of M. marinum, avirulent M. marinum, isolated nucleic acid comprising oligonucleotide of SEQ ID NO: 4, 6, 8, 11, 13, 21, 23, 25, 27, 29, 31, 39, 41 and nucleic acids which is complementary to or which can hybridize under conditions of high stringency to a portion of said nucleic acid identified by said SEQ ID NOs, pharmaceutical composition comprising avirulent M. marinum bacterium, attenuated M. marinum vaccine comprising said avirulent bacterium, method for isolating a mutagenized M. marinum bacterium which exhibits reduced virulence in a host.

2. Claims: 6-27, 43-45, 47

Method for identifying a virulence gene of M. tuberculosis, method for generating avirulent M. tuberculosis bacterium, avirulent M. tuberculosis comprising one or more mutated genes according to claims 9-27, pharmaceutical composition comprising avirulent M. tuberculosis bacterium, attenuated M. tuberculosis vaccine comprising said avirulent bacterium.

3. Claim : 48

An isolated polyketide made by the M. marinum polyketide synthase gene.

4. Claim : 49

An isolated polyketide made by the M. tuberculosis polyketide synthase gene.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 00/25512

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9909186 A	25-02-1999	FR 2767336 A	19-02-1999
		FR 2767337 A	19-02-1999
		AU 9076598 A	08-03-1999
		EP 1003888 A	31-05-2000